PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 189

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W.1

3 June 1947

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
(Cambridge, and Bentley House, London)

Agents for Canada and India: Macmillan

CONTENTS

SERIES A VOLUME 189

No. A 1016. 27 March 1947

PAGE

Interaction of steam and charcoal at low pressures. By R. F. Strickland-Constable 1

On the dimension of normal spaces. By P. Alexandroff 11

The crystal structures of two polyamides ('nylons'). By C. W. Bunn and E. V. Garner. (Plates 1 and 2) 39

Limitations of the Newtonian time scale in relation to non-equilibrium rheological states and a theory of quasi-properties. By G. W. S. Blair, B. C. Veinoglou and J. E. Caffyn 69

Kinetic boundary friction. By J. R. Bristow. (Plate 3) 88

A general kinetic theory of liquids. II. Equilibrium properties. By H. S. Green 103

Tables of second virial and low-pressure Joule-Thomson coefficients for intermolecular potentials with exponential repulsion. By R. A. Buckingham and J. Corner 118

The oblique reflexion of very long wireless waves from the ionosphere. By M. V. Wilkes 130

No. A 1017. 17 April 1947

The fine structure of the spectral lines of H_2: measurements with a reflexion echelon. By E. W. Foster and Sir Owen Richardson, F.R.S. (Plate 4) . 149

The fine structure of the lines of the H_2 spectrum which go down to the $2p\,{}^3\Pi$ states. By E. W. Foster and Sir Owen Richardson, F.R.S. 175

Radiative equilibrium in the ionosphere. By R. v. d. R. Woolley 218

Atmospheric tides in the ionosphere. I. Solar tides in the F_2 region. By D. F. Martyn 241

Temperature changes in the crystal structure of barium titanium oxide. By H. D. Megaw. (Plate 5) 261

Visual adaptation in relation to brief conditioning stimuli. By B. H. Crawford. (Abstract) 284

No. A 1018. 1 May 1947

PAGE

Effect of crossed electric and magnetic fields on the helium spectrum. By J. S. Foster, F.R.S. and E. R. Pounder. (Plate 6) 287

The surprising amount of energy which can be collected from gases after the electric discharge has passed. By Lord Rayleigh, F.R.S. 296

Fragmentation of shell cases. By N. F. Mott, F.R.S. 300

On a method of increasing contrast in microscopy. By Sir Thomas Merton, Treas.R.S. (Plate 7) 309

The spectra of flames containing oxides of sulphur. By A. G. Gaydon and G. Whittingham. (Plates 8 and 9) 313

On the simple group of order 25920. By J. A. Todd 326

The inversion spectrum of ammonia at centimetre wave-lengths. By B. Bleaney and R. P. Penrose 358

Infinite irreducible representations of the Lorentz group. By Harish-Chandra . 372

Thermal diffusion in mixtures of the inert gases. By K. E. Grew 402

A theory of the dependence of the rate of detonation of solid explosives on the diameter of the charge. By H. Jones 415

No. A 1019. 3 June 1947

Antoine Laurent Lavoisier, 26 August 1743—8 May 1794. By Sir Harold Hartley, F.R.S. (Plates 10 and 11) 427

Determination of the velocity coefficients for polymerization processes. I. The direct photopolymerization of vinyl acetate. By G. M. Burnett and H. W. Melville, F.R.S. 456

Determination of the velocity coefficients for polymerization processes. II. The photosensitized polymerization of vinyl acetate by benzoyl peroxide. By G. M. Burnett and H. W. Melville, F.R.S. 481

Determination of the velocity coefficients for polymerization processes. III. Effect of environment on polymerization processes. By G. M. Burnett and H. W. Melville, F.R.S. 494

The dehydration of ammonium, potassium and some mixed alums. By G. P. Acock, W. E. Garner, F.R.S., J. Milsted and H. J. Willavoys. (Plates 12 to 14) 508

Slow neutron-induced activity in gold. By K. Siegbahn 527

Mixed boundary conditions in the relaxational treatment of biharmonic problems (plain strain or stress). By L. Fox 535

Vertical transport of heat by turbulence in the atmosphere. By C. H. B. Priestley and W. C. Swinbank 543

Index 563

Interaction of steam and charcoal at low pressures

By R. F. Strickland-Constable, D.Phil.

(*Communicated by C. N. Hinshelwood, F.R.S.—Received* 16 *April* 1946)

The reaction between charcoal and water vapour at temperatures up to 700° C and at pressures of less than 1 mm. has been studied.

Water vapour, hydrogen and carbon monoxide have all been shown to be reversibly adsorbed at the reaction temperatures. Hydrogen is adsorbed strongly and rapidly; carbon monoxide is adsorbed in considerably smaller quantity and much more slowly.

When charcoal reacts with water vapour at low pressures, the pressure falls during the rapid initial adsorption of the water vapour, and subsequently rises during the desorption of the product gases. The proportions of carbon monoxide and hydrogen in the product gases were found to vary throughout the reaction in a special manner. The variations can, however, be simply explained on the basis of the adsorption data.

The kinetics of the reaction at higher pressures are discussed: they can likewise be adequately explained on the basis of the adsorption data.

Reasons are given for supposing that the water vapour is adsorbed in the molecular condition.

The course of gas/solid reactions is generally explained as occurring in an adsorbed layer on the surface of the solid. The nature and extent of the adsorption of the reactant and product gases is often deduced from the course of the reaction without direct measurement. In the present research, by the use of low gas pressures and a solid with a large surface, an attempt has been made to measure the adsorption of the gases separately, and to relate these to the course of the reaction.

Experimental

The apparatus is illustrated in figure 1. It consists of: *A*, the reaction bulb (vol. 185 ml.); *B*, the capacity bulb (vol. 940 ml.), which was used to measure out charges of water vapour; *C*, connexion to the Macleod gauge; *D*, tap to reaction bulb; *E*, connexion leading to the analysis bulb; *F*, tap leading to gas storage apparatus; *G*, a thermocouple; *H*, furnace; *K*, tap leading to vacuum pumps.

The reaction bulb, of porcelain, was heated in an electric furnace whose temperature was measured by means of the thermo-couple. In the storage apparatus water and the various gases required were stored in a series of bulbs. The apparatus, except for the reaction bulb, was of soda glass. The usual precautions with regard to cleanliness that are necessary in low-pressure technique were observed. The tap *K* led to a Gaede mercury-vapour pump, by which the apparatus could be rapidly evacuated.

To measure the pressure of water vapour the compression stem of the Macleod gauge was heated in a hot-water bath at 100° C. The latter consisted of an electrically heated glass jacket containing water to which a small amount of soap had been added to facilitate steady boiling. Thus as the water vapour which was trapped at room temperature in the body of the Macleod was compressed into the compression stem, it was raised to 100° C, and was thus maintained above its dew-point. When used in this way with permanent gases the readings were found to be increased in the calculated ratio as compared with those read on an unheated instrument. For use with water vapour the instrument was calibrated against known pressures which were established by surrounding a portion of the main apparatus containing water by a bath held at controlled low temperatures. In this way it was shown that over the range of 150 to 1400 $\times 10^{-3}$ mm. the pressure read was fairly accurately equal to 90 % of the value obtained with permanent gases. This correction was accordingly made to all readings of pure water-vapour pressures.

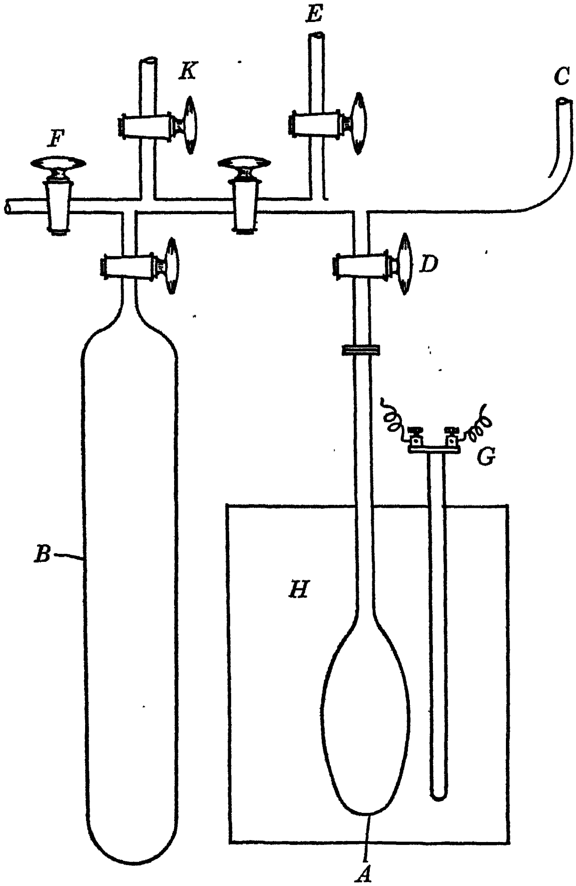


Figure 1

In previous work with similar apparatus (Strickland-Constable 1944) and using gas pressures ranging from 0·02 to 1·0 mm., it had been found that all the gases used, including N_2O and CO_2, obeyed Boyle's law: thus if one vessel at a known pressure was put into communication with a second evacuated vessel, the proportional change in pressure was independent of the initial pressure in the first vessel, and of the nature of the gas. But this was far from the case with water vapour, of which a large proportion appeared to reside on the surface of the glass, and not in the gas phase.

Many experiments were performed to investigate the behaviour of water vapour, and as the result of these and of the work of Frank (1929), the following conclusions were found to apply to an apparatus of the dimensions used in the present work:

When the measured pressure is very low the greater part of the water vapour present resides on the surface of the vessel, owing to the strong adsorption of the first one or two layers of molecules. On increasing the pressure the proportion present in the gas phase increases, and over about 300×10^{-3} mm. at least half the total water is so present. As, however, the pressure approaches saturation the situation is again reversed, and an ever greater proportion resides on the glass surface. Two sets of experiments can be briefly described to illustrate these points. Hydrogen and oxygen were introduced into the analysis bulb, their pressures being measured on the Macleod. On heating the platinum filament, water vapour equal in pressure to the figures given in column 1 of table 1 should have been produced: the pressures actually observed are shown in column 2. Column 3 gives the observed as a percentage of the expected values.

TABLE 1. PRESSURES IN MM. $\times 10^{-3}$

expected	found	found/expected %
560	350	62½
408	220	54
200	75	37
30	0	0

In another set of experiments a rather high pressure of water vapour was established in a small vessel with a high surface-volume ratio, and the water vapour was then expanded into the analysis bulb. When 10 mm. pressure was established in the small vessel, the pressure developed on expansion into the analysis bulb was 325×10^{-3} mm., although with a permanent gas the pressure would have been only 108×10^{-3} mm. The temperature corresponded to a saturation pressure of about 13 mm. A layer of water at least 20 mol. thick on the walls of the small vessel was calculated to be necessary to account for the above result.

In what follows the quantities of water vapour used are recorded by pressure measurements, though it must be remembered throughout that the actual quantities of water present are usually more than double those corresponding to the recorded pressures, and that the total amounts present are not even directly proportional to these pressures.

The analysis bulb and the method of using it for analyzing mixtures of CO, CO_2, and O_2 have been described in a previous paper by the author (1944). In the present work the mixture could consist of H_2O, H_2, CO_2 and CO. Although it was useless to attempt to determine accurately the amount of H_2O present, owing to the strong adsorption on the walls of the apparatus, it was sometimes possible, by using the P_2O_5 reagent tube, to obtain a minimum value. Usually, however, H_2O was first absorbed in the P_2O_5, and the analysis then carried out on the dry mixture remaining as follows:

CO_2 was first determined by absorption in the Sofnolite. The Sofnolite cut-off was then closed (the P_2O_5 cut-off being left open throughout the analysis), oxygen was added, and the H_2 and CO were burned on the platinum filament. The fall in pressure was noted, the Sofnolite tube was then opened to absorb the CO_2 formed, and a second fall in pressure was noted. These two pressure drops allow of the calculation of the amounts of CO and H_2 present. The method was found to be accurate at pressures of over 200×10^{-3} mm. At lower pressures H_2 tended to be too low and CO too high. A sample test analysis is as follows:

total pressure		400×10^{-3} mm.
took	CO 48 %	H_2 52 %
found	CO 46 %	H_2 51 %

It was established that at the temperature at which the platinum filament was heated for the combustion of the H_2 and CO, no burning of CH_4 took place, so that the latter gas would have remained as an unburnt residue in the analysis. No methane was in fact ever found.

The charcoal consisted of 5 g. of a steam-activated coconut charcoal known as Dorsite, as used in previous experiments (Strickland-Constable 1938).

RESULTS

Before giving the detailed experimental results, an outline of the general theory of the reaction which was deduced from them will be given. Most of the experiments were performed at 700° C, and this temperature will be assumed unless the contrary is specified. It appears that even at this temperature H_2O is reversibly adsorbed on the surface of charcoal, and this adsorption takes place very rapidly. In the adsorbed condition the water reacts with the charcoal to form CO and H_2, and these products are then desorbed. Both CO and H_2 are themselves reversibly adsorbed at this temperature.

When therefore a low pressure of water vapour is allowed to react with a charcoal previously outgassed at a much higher temperature, the water is adsorbed and the pressure falls at once to a very low figure where it remains indefinitely. The water in the absorbed condition reacts to form CO and H_2, but these are likewise retained by strong adsorption on the surface. Only by repeating this process a number of times or by using a comparatively high pressure of water vapour does the charcoal become more or less saturated with product gases, and the sequence of events, on introducing a further charge of water vapour, is as follows:

The pressure falls at first to a rather low value and then begins to rise again. The initial fall is due to adsorption of water, and the subsequent rise to desorption of product. If a sample of the gas phase is taken early in the reaction, when only a small amount of product has been evolved, it is found to contain a much higher proportion of H_2 than a sample taken later on, which consists largely of CO. This was one of the principal features of the reaction. The explanation appears to depend on the fact that at equilibrium H_2 is adsorbed much more strongly than CO; at the same time the adsorption and desorption of H_2 are far more rapid than is the case with CO, so that H_2 adsorption equilibrium is established very quickly, that of CO much more slowly. In the early stages of the reaction the H_2 will desorb more rapidly and hence be in excess in the gas phase. Later in the reaction, as adsorption equilibrium is approached, CO will be in excess in the gas phase, owing to its much smaller equilibrium adsorption.

In the first set of experiments to be described the charcoal was outgassed at 1000° C, and thereafter a series of runs was carried out. In these runs a charge of water vapour was admitted to the capacity bulb and its pressure measured on the Macleod. Tap *D* was then opened so that the capacity bulb and reaction bulb were placed in communication, and the change of pressure with time was observed. The pressure at $t = 0$ was estimated from the original pressure in the capacity bulb, this estimate being based on sharing experiments previously carried out at the appropriate temperature when the reaction bulb contained no charcoal. The results are given in table 2.

TABLE 2. CAPACITY BULB METHOD, 700° C

	pressures (mm. × 10^{-3}) at t (min.)						H_2O adsorbed ml. of vapour
run	0	1	2	6	15	30	at N.T.P.
118	990	360	290	130	70	70	1·440
119	930	450	390	260	170	160	1·270
120	1170	720	640	590	510	470	1·400

In the first run the initial pressure was 990×10^{-3} mm. The pressure fell rapidly however, and after 30 min. only about 70×10^{-3} mm. were left. The residual gas consisted of H_2 and CO with little or no water vapour. The 5 g. of charcoal had therefore adsorbed 1·44 ml. of water vapour (nominally calculated to N.T.P.) with the formation of little or no gaseous product. The object of these runs was to saturate the charcoal, since unless this was done no substantial amount of product gases could be obtained. Two more runs were accordingly carried out by the same method, each run being preceded by a short evacuation to remove gases left in the gas phase. At the end of the third run the gas pressure was about six times that after the first, showing progressive saturation of the charcoal. Finally, before proceeding to the next series, a pressure of 16 mm. of water vapour was established in the bulb and maintained there for 20 min.

In the next series of runs the capacity bulb was not used, the H_2O being introduced into the Macleod alone, where its pressure was measured, before opening the tap to the reaction vessel. The volume of Macleod + reaction bulb (the latter reduced to room temperature) was only 200 ml. as compared with a total volume of 1140 ml.

when using the capacity-bulb method. After opening tap *D* the course of the reaction was followed as before, and the results of some runs are given in table 3.

TABLE 3. NORMAL METHOD, 700° C

run	pressure (mm. × 10^{-3}) at t (min.)					
	0	1	2	10	20	30
126	750	260	280	390	—	—
127	660	230	240	330	400	—
129	820	300	310	480	600	690

The apparatus was evacuated between runs for sufficiently long to reduce the rate of gas evolution to a conveniently low value: 2 or 3 min. was usually sufficient. It will be seen from the table that the pressure falls during the first minute to approximately one-third the initial value, and then rises slowly. The initial fall represents the very rapid adsorption of the water vapour, and the slow subsequent rise the evolution of the products of reaction. The reaction $H_2O + C = CO + H_2$ demands a doubling of the initial pressure. As will be seen from the table, however, the pressures never even regain their initial values. This appeared to be due to the fact that although the rate of desorption was high at very low pressures, it fell off rapidly as the pressure rose. Thus a considerable part of the gas produced in each run was only desorbed during the evacuation of the bulb between runs.

In other runs by the same method, samples of gas were taken for analysis (see table 4).

TABLE 4. NORMAL METHOD, 700° C

run	pressure at $t=0$	time of sampling (min.)	analysis %			
			CO_2	H_2	CO	unknown
128	660	1	12	42	44	2
130	700	½	13	47½	36	3½
127	540	21	3½	12½	84½	−½
129	693	34	4	18	78½	−½

The time of sampling appeared to have an important effect on the composition of the sample. Early samples had a higher proportion of H_2 compared to later samples, as will be clearly seen in the above table. The explanation of this observation has already been given. (It will, of course, be noticed that the CO_2 content of the early samples is higher than that of the later samples. The effect is, however, insufficiently great to account for the much larger changes in hydrogen concentration.)

In yet another type of experiment the water container was allowed to communicate with the reaction bulb until a high pressure of vapour had been established. In run 199 for instance a pressure of 18 mm. of water vapour was established in the bulb. A sample was taken after 18 min., and, after removal of water vapour, the dry gases analyzed to: CO_2 23 %, H_2 24 %, CO 51 %. The bulb was then evacuated for 1 min., and a sample of gas allowed to accumulate for ½ min.; the dry gases in this case analyzed to: CO_2 2 %, H_2 93 %, CO 7½ %. The explanation of the change in hydrogen content is the same as before: the sample taken 18 min. after the introduction of the water vapour corresponds to the later samples of table 3, and had a low hydrogen content owing to approach to adsorption equilibrium. The sample taken immediately after a short evacuation consisted of nearly pure hydrogen owing to the high rate of desorption of the latter, as compared with CO.

That this view was the correct one was confirmed by the following further experiments. Equal volumes of CO and H_2 were admitted to the reaction bulb at 700° C and left for over 2 hr., at the end of which time the pressure was 11½ mm. A sample of the gas phase now analyzed to: CO_2 0 %, H_2 8 %, CO 81 %, N_2 11 %. The bulb was then evacuated for 1½ min., and a further sample was allowed to collect during 2¼ min., which gave on analysis: CO_2 1 %, CO 8 %, H_2 87 %. The 11 % of N_2 recorded in the first analysis was attributed to the presence of a small amount of this gas as an impurity in the CO used. Owing to its not being adsorbed the small proportion present in the initial gas became concentrated in the gas phase as a substantial constituent.

The difference in the rates of adsorption of CO and H_2 was shown directly in an experiment in which a charge of CO was admitted to the bulb at 700° C after the charcoal had been outgassed at 1000° C. The variation of pressure with time is shown in table 5.

TABLE 5. CO, 700° C

t (min.)	0	¼	2	5	15	30
pressure (mm. $\times 10^{-3}$)	3620	3500	3260	3020	2560	2040

After evacuation for several minutes, to remove all the gaseous CO, a charge of H_2 was admitted (see table 6).

TABLE 6. H_2, 700° C

t (min.)	0	¼	2	5	15
pressure (mm. $\times 10^{-3}$)	4600	1720	342	246	213

Thus H_2 appears to be adsorbed far more quickly than CO: in the first 2 min. 360×10^{-3} mm. of CO were adsorbed, as against 4258×10^{-3} mm. of H_2, and this in spite of the fact that the H_2 adsorption was taking place on a charcoal already partly covered with CO.

The difference in the total *quantities* which could be adsorbed was shown in a qualitative way in the following experiments: CO was admitted to the reaction bulb, which was maintained at 700° C for half an hour, during which time a fall in pressure of 2400×10^{-3} mm. was recorded. The bulb was then evacuated for 7 sec.; the pressure was now found to be 270×10^{-3} mm., and it continued to rise slowly. Thus after adsorbing 2400×10^{-3} mm. of CO the equilibrium pressure is over 270×10^{-3}. (It may be noted that 1 mm. pressure in the Macleod + reaction bulb at 700° C is equivalent to about 0·25 ml. of gas at N.T.P.)

In a similar experiment with H_2, a pressure of 8600×10^{-3} mm. was adsorbed, after which the equilibrium pressure was *less* than 200×10^{-3} mm. This experiment shows that the adsorption of H_2 is at least three times that of CO, and probably considerably more.

Experiments were made to show that the adsorption of water vapour is reversible. At 400° C a pressure of 18 mm. of water vapour was established in the bulb and maintained for 15 min. The bulb was then evacuated for 5 min., a time sufficient to remove very completely any gas originally present in the gas phase. During the following 30 min. a pressure of about 0.5 mm. of gas accumulated in the bulb, and this was shown to consist of over 95 % water vapour.

A very similar run at 600° C gave a gas sample containing over 40 % of water vapour. At this temperature the rate of reaction, although much slower than at 700° C, was nevertheless beginning to be appreciable. At 700° C, owing to the much greater rate of reaction, it was more difficult to obtain evidence, but even at this temperature a minimum of 9 % of water vapour was found in a sample of gas which was allowed to collect after 2 min. evacuation following a saturation of the charcoal as before.

The presence of certain amounts of CO_2 has been disclosed in the various analyses quoted. In table 4 the two early samples had a ratio of CO_2 to CO of 1 to 3·7, whereas the two late samples had a ratio of less than 1 in 20. In run 199 the proportion of CO_2 to CO was 1 to 2·2. No special study was made of the conditions of formation of CO_2, nor was the adsorption measured. It is most probably formed as the result of the secondary reaction $CO + H_2O = CO_2 + H_2$. The amounts formed were in no case sufficient to obscure the observations that were being made on the main reaction.

DISCUSSION OF RESULTS

A general explanation of the low-pressure reaction in terms of the measured adsorption data has already been given at the beginning of the paragraph headed 'Results', and need not be repeated here.

In what follows, the word adsorption will be used for a strictly reversible process, which will be clearly distinguished from an irreversible sorption. It has been shown that both the products, H_2 and CO, and the reactants, water vapour, are reversibly adsorbed up to at least 700° C. The fact that the sorption of water vapour is reversible is of interest from the point of view of the mechanism of sorption at these relatively high temperatures. Sorption at these temperatures is often considered to involve dissociation of the molecule into atoms: if, however, the sorption of H_2O is reversible such a dissociation would in this case seem to be most unlikely in view of the known very high rates of reaction of such gases as O_2 and N_2O with charcoal at these temperatures. It would appear, in fact, that forces must exist which hold the complete molecules H_2O, CO and H_2 on the surface in such a manner that they can easily be desorbed again into the gas phase.

In considering the bearing of the present results on the steam-carbon reaction at higher pressures reference is made to the comprehensive work of Gadsby, Sykes & Hinshelwood (1946). In this work, which was carried out at steam pressures of from 100 to 760 mm., it is shown that the rate of reaction is given by an expression of the form of

$$\text{rate} = \frac{k_a p_1}{1 + k_b p_2 + k_c p_1}, \tag{1}$$

where p_1 = partial pressures of H_2O, and p_2 = partial pressures of H_2. This formula corresponds to a reaction that at high steam pressures approaches zero order with regard to steam, and is retarded by hydrogen.

Now such a reaction formula can easily be accounted for on the basis of the present results. One of the principal points established here is the slowness of the desorption of the CO, compared with other stages of the reaction. Thus at 700° C the desorption of CO was found to take many minutes, while the adsorption of H_2O was largely completed in a fraction of a minute. Qualitatively, therefore, the reaction may be expected to be controlled by the slow desorption of CO and to approach zero order at higher pressures. The strong adsorption of H_2, which was demonstrated, would account for the retardation by H_2, although to complete the evidence for this it would be desirable to show that the adsorption of H_2 reduces the amount of H_2O which can be adsorbed.

Quantitatively it is possible to deduce a formula for reaction rate on the basis of such a mechanism, assuming, that is to say, reversible adsorption of H_2O and H_2, reaction of adsorbed H_2O to adsorbed CO, followed by slow desorption of CO (owing to its weak adsorption the reverse reaction consisting of the adsorption of gaseous CO can be neglected). Let

	H_2O	H_2	CO
fractions of surface covered be	s_1	s_2	s_3
partial pressures of gases be	p_1	p_2	—

Then rate of desorption of CO = $k_6 s_3$, and in the steady state this must equal the rate of formation of CO from adsorbed H_2O, which equals $k_5 s_1$ (k_5, k_6, etc., being appropriate constants).

Therefore
$$k_6 s_3 = k_5 s_1. \tag{2}$$

For H_2O, rate of adsorption = rate of desorption + rate of reaction to CO:

$$k_1 p_1 (1 - s_1 - s_2 - s_3) = k_2 s_1 + k_5 s_1. \tag{3}$$

For H_2 adsorption equilibrium

$$k_3 p_2 (1 - s_1 - s_2 - s_3) = k_4 s_2. \tag{4}$$

Equations (2), (3) and (4) can be solved to give

$$\text{rate of reaction} = k_6 s_3 = \frac{\dfrac{k_1 k_5}{k_2 + k_5} p_1}{1 + \dfrac{k_3}{k_4} p_2 + \dfrac{k_1 (1 + k_5/k_6)}{k_2 + k_5} p_1}. \tag{5}$$

This equation is of the same form as (1). It is also nearly the same as a formula deduced by Gadsby *et al.* (1946), which does not, however, distinguish the separate steps of the reaction of adsorbed H_2O and the desorption of adsorbed CO. This formula gives the rate of reaction as

$$\frac{\dfrac{k_1 k_5}{k_2+k_5} p_1}{1+\dfrac{k_3}{k_4} p_2+\dfrac{k_1}{k_2+k_5} p_1}. \tag{6}$$

Equations (5) and (6) differ only in the last term of the denominator where an additional factor $(1+k_5/k_6)$ appears in (5). This whole term is the one which is responsible for the reaction tending to zero order at high steam pressures: and if this effect is due to a relatively slow desorption of CO, then k_5/k_6 will be large compared with unity, and will determine the importance of the term. If, on the other hand, the rate of desorption of CO were fast, then k_5/k_6 would be small and the size of the term would be determined by the ratio of k_1 to k_2+k_5; in this case the zero order at high pressures would be conditioned by a slow rate of desorption of H_2O (k_2 small), together with a slow rate of reaction of adsorbed H_2O (k_5 small), or in other words by the surface being largely covered by adsorbed H_2O.

It is considered that the present results favour the view that the slow desorption of CO is the controlling factor, and that the mechanism will therefore be represented by equation (5) with a value of k_5/k_6 that is large compared with unity.

The writer wishes to express his special thanks to Professor C. N. Hinshelwood for his advice and help throughout the work described in this paper. The work was carried out as part of the programme of the Fuel Research Board of the Department of Scientific and Industrial Research, and is published by permission of the department.

REFERENCES

Frank 1929 *J. Phys. Chem.* **33**, 970.
Gadsby, Sykes & Hinshelwood 1946 *Proc. Roy. Soc.* A, **187**, 129.
Strickland-Constable 1938 *Trans. Faraday Soc.* **34**, 1375.
Strickland-Constable 1944 *Trans. Faraday Soc.* **40**, 333.

On the dimension of normal spaces

BY P. ALEXANDROFF

Mathematical Institute of the Academy of Sciences, Moscow

(*Communicated by M. H. A. Newman, F.R.S.—Received* 24 *April* 1946)

The main purpose of this paper is to generalize for the case of normal spaces the fundamental theorem of the homological dimension theory and the existence theorem on Cantor manifolds in n-dimensional spaces.

This generalization is based on

(1) The generalization† of the same theorems, and many others, for bicompacta (i.e. bicompact Hausdorff spaces)—§§ 4, 5.

(2) The study of certain 'maximal' subspectra of the spectrum of a bicompactum—§ 3.

(3) A new approach to the maximal, or Čech, extension of a normal space—§ 6.

(4) A suitable definition of the generalized notion of Cantor manifold, which is combinatorial in character and, when applied to compacta, corresponds to the classical definition given by Urysohn—§ 7.

In §§ 1, 2 I recall well-known definitions regarding homological invariants of bicompacta, and their coverings and spectra (see, for example, my papers (1941 *a*, 1943) and the fundamental work of Lefschetz (1942, chapter VI)).

Notations. $\cup$ denotes set-theoretical sums, $\cap$ intersections of sets; Λ stands for the void set.

1. SIMPLICIAL SPECTRA

A simplicial spectrum (as defined in Alexandroff 1943, § 28) is a directed set (Lefschetz 1942, p. 4) $C = \{C_\alpha, \varpi_\alpha^\beta\}$ of finite simplicial complexes (or 'nerves'—see Alexandroff 1943, §§ 28, 30) C_α, together with certain simplicial mappings ('projections') ϖ_α^β, under the following conditions:

(1) If C_β follows C_α in the directed set C ($\beta > \alpha$ is written for this), then there is given a finite number $\geqslant 1$ of projections ϖ_α^β of C_β in C_α.

(2) If $\varpi_\alpha'^\beta$ and $\varpi_\alpha''^\beta$ are two projections (of C_β in C_α), then for each simplex $t_\beta \in C_\beta$ the simplices $t'_\alpha = \varpi_\alpha'^\beta t_\beta$ and $t''_\alpha = \varpi_\alpha''^\beta t_\beta$ are faces of a certain simplex $t_\alpha \in C_\alpha$ ($t'_\alpha \leqslant t_\alpha$, $t''_\alpha \leqslant t_\alpha$ is written for this, thus considering each simplex as its own 'improper' face).

(3) If $\gamma > \beta > \alpha$ and ϖ_β^γ, ϖ_α^β are projections, then the mapping $\varpi_\alpha^\beta \varpi_\beta^\gamma$ of C_γ in C_α is a projection.

---

† These generalizations could be obtained entirely on the basis of the methods developed in Alexandroff (1932, 1940, 1941 *a*) and Alexandroff, Hopf & Pontrjagin (1935); but a shorter and much more elegant way is given by combination of these methods with those used by Hurewicz & Wallman (1941) in their excellent book. It is, however, necessary to recall, since no reference is made by Hurewicz & Wallman, that the main theorems of the homological theory of dimension were established in the paper by Alexandroff (1932).

The oriented simplices of C_α will still be denoted by t_α, the group of all r-dimensional chains x_α of C_α (with respect to a fixed coefficient-group) and the Betti groups (the homology—or Δ—and the 'cohomology' or ∇—groups) will be denoted by L^r_α and Δ^r_α, ∇^r_α respectively.

Then any projection ϖ^β_α generates the homomorphism

ϖ^β_α of L^r_β into L^r_α	π^α_β of L^r_α into L^r_β

determined by

$(\varpi^\beta_\alpha x^r_\beta . t^r_{\alpha i}) = \sum\limits_j \epsilon_{ij}(x^r_\beta . t^r_{\beta j})$ with $\epsilon_{ij} = 1, -1, 0$ according to $\varpi^\beta_\alpha t^r_{\beta j} = t_{\alpha i}, = -t_{\alpha i}, \neq \pm t_{\alpha i}$	$(\pi^\alpha_\beta x^r_\alpha . t^r_{\beta j}) = (x^r_\alpha . \varpi^\beta_\alpha t^r_{\beta j})$

(see e.g. Alexandroff 1941*a*, § 2). The operators ϖ^β_α and π^α_β are conjugate homomorphisms and commute with the boundary operators Δ and ∇ respectively; they produce therefore homomorphisms ϖ^β_α, π^α_β of the Betti groups, namely,

the homomorphism ϖ^β_α of Δ^r_β into Δ^r_α	the homomorphism π^α_β of ∇^r_α into ∇^r_β.

The condition (2) in the definition of a simplicial spectrum ensures that these homomorphisms do not depend on the special choice of the projection ϖ^β_α of C_β in C_α. Thus we have the

inverse group spectrum $\{\Delta^r_\alpha, \varpi^\beta_\alpha\}$	the direct group spectrum $\{\nabla^r_\alpha, \pi^\alpha_\beta\}$

(see Alexandroff 1941*a*, §§ 3, 4; Alexandroff 1943, §§ 22, 29; or Lefschetz 1942, chapter VI). The corresponding limit groups are the r-dimensional Betti groups Δ^r_C, ∇^r_C of the simplicial spectrum C with respect to the given coefficient-group (discrete for ∇ and bicompact for Δ).

The simplest direct definition of the group ∇^r_C is the following:

Take the set of all r-dimensional ∇-cycles z^r_α of all $C_\alpha \in C$; the ∇-cycles z^r_α and z^r_β belong by definition to the same (r-dimensional) ∇-class z^r, if there exists C_γ following in C on both C_α and C_β, such that

$$\pi^\alpha_\gamma z^r_\alpha \sim \pi^\beta_\gamma z^r_\beta \quad \text{in} \quad C_\gamma.$$

The r-dimensional ∇-classes are by definition the elements of ∇^r_C. To get the sum of two ∇-classes z'^r and z''^r take arbitrarily $z^r_\alpha \in z'^r$, $z^r_\beta \in z''^r$, and a C_γ following in C on both C_α and C_β; the ∇-class z^r containing

$$\pi^\alpha_\gamma z'^r_\alpha + \pi^\beta_\gamma z''^r_\beta$$

is by definition $z'^r + z''^r$.

Now suppose that in each C_α a closed subcomplex A_α is given. Denote the Betti groups of A_α by $\Delta^r_{\alpha 0}$, $\nabla^r_{\alpha 0}$. Suppose that for any pair $\beta > \alpha$ and any choice of the projection ϖ^β_α it is given that $\varpi^\beta_\alpha A_\beta \subseteq A_\alpha$. Suppose, furthermore, that for $t_\beta \in A_\beta$

the condition (2) in the definition of a simplicial spectrum always holds with $t_\alpha \in A_\alpha$. Then the A_α with the mappings ϖ^β_α considered on them form a spectrum, which is called a subspectrum $A = \{A_\alpha, \varpi^\beta_\alpha\}$ of the spectrum $C = \{C_\alpha, \varpi^\beta_\alpha\}$. Denote the open subcomplex $C_\alpha \setminus A_\alpha$ by H_α; this subcomplex (more precisely, the system of its oriented simplices) is a cell-complex in the sense of Kolmogoroff and Tucker (see e.g. Alexandroff 1943, §§5–7), its Betti groups will be denoted by $\Delta^r_{\alpha 1}$, $\nabla^r_{\alpha 1}$. Now consider any chain $x^r_{\alpha 1}$ on H_α with coefficients belonging to a bicompact group. The set of all r-chains x^r_α on C_α which coincide with $x^r_{\alpha 1}$ on H_α is denoted by $E^{\alpha 1}_\alpha x^r_{\alpha 1}$ and called the Δ-extension of $x^r_{\alpha 1}$ over C_α (this slight deviation from the definitions of Alexandroff (1943) corresponds to the real state of affairs and is thus reasonable). For a chain x^r_α on H_α with respect to a discrete coefficient-group the trivial ∇-extension $\mathscr{E}^{\alpha 1}_\alpha x^r_{\alpha 1}$ over C_α is defined as the chain x^r_α on C_α equal to $x^r_{\alpha 1}$ on H_α and equal to zero on $A_\alpha = C_\alpha \setminus H_\alpha$.

The intersection-operator of Alexandroff (1943, §11) is still denoted by $J^{\alpha 1}_\alpha$; for any chain x^r_α on C_α, and any system of chains x^r_α which are equal to each other on H_α the chain $J^\alpha_{\alpha 1} x^r_\alpha$ is by definition that chain on H_α which is equal there to x^r_α.

We define for any $z^r_{\beta 1} \in \zeta^r_{\beta 1} \in \Delta^r_{\beta 1}$

$$\varpi^{\beta 1}_{\alpha 1} z^r_{\beta 1} = J^\alpha_{\alpha 1} \varpi^\beta_\alpha E^{\beta 1}_\beta z^r_{\beta 1}.$$

This gives a homomorphism $\varpi^{\beta 1}_{\alpha 1}$ of $\Delta^r_{\beta 1}$ in $\Delta^r_{\alpha 1}$ and therefore an inverse group spectrum $\{\Delta^r_{\alpha 1}, \varpi^{\beta 1}_{\alpha 1}\}$ with the bicompact limit group denoted by Δ^r_H and called the r-dimensional Δ-group of C modulo A.

In a dual way consider for $z^r_{\alpha 1} \in \mathfrak{z}^r_{\alpha 1} \in \nabla^r_{\alpha 1}$ the ∇-cycle $\pi^\alpha_\beta \mathscr{E}^{\alpha 1}_\alpha z^r_{\alpha 1}$ which is zero on A_β so that $J^\beta_{\beta 1} \pi^\alpha_\beta \mathscr{E}^{\alpha 1}_\alpha z^r_{\alpha 1}$ is a ∇-cycle on H_β. Define

$$\pi^{\alpha 1}_{\beta 1} z^r_{\alpha 1} = J^\beta_{\beta 1} \pi^\alpha_\beta \mathscr{E}^{\alpha 1}_\alpha z^r_{\alpha 1}.$$

This gives a homomorphism $\pi^{\alpha 1}_{\beta 1}$ of $\nabla^r_{\alpha 1}$ into $\nabla^r_{\beta 1}$ (conjugate to the homomorphism $\varpi^{\beta 1}_{\alpha 1}$—see Alexandroff 1943, §§20–22 and 29). In this way the direct spectrum $\{\nabla^r_{\alpha 1}, \pi^{\alpha 1}_{\beta 1}\}$ is obtained with the discrete limit group

$$\nabla^r_H = \lim\, (\nabla^r_{\alpha 1}, \pi^{\alpha 1}_{\beta 1})$$

called the r-dimensional ∇-group of C modulo A. To repeat once more: the elements of ∇^r_H are ∇-classes, two ∇-cycles $z^r_{\alpha 1}$ and $z^r_{\beta 1}$ belonging by definition to the same ∇-class if there exists a C_γ, $\gamma > \beta$, $\gamma > \alpha$ such that

$$\pi^{\alpha 1}_{\gamma 1} z^r_{\alpha 1} \sim \pi^{\beta 1}_{\gamma 1} z^r_{\beta 1} \quad \text{on} \quad H_{\gamma 1}.$$

REMARK 1. A simplicial spectrum is called *exact* if for each pair $\beta > \alpha$ there is but one projection ϖ^β_α and this projection maps C_β on to C_α.

REMARK 2. Let Ω, Ω' be two directed sets; suppose that each element of Ω' is an element of Ω and that $\beta > \alpha$ in Ω' implies $\beta > \alpha$ in Ω. Then Ω' is called a directed subset of Ω. A directed subset Ω' of Ω is called a *confinal subset of the directed set* Ω if for any element α or Ω there is an element β' of Ω' which follows on α in Ω.

2. CANONICAL SETS AND SPECTRA OF A SPACE

Throughout this paper 'space', 'mapping of a space', 'covering' mean *normal* space, *continuous* mapping, *finite* covering, respectively.

Let R be a space, M an arbitrary point set in R. We denote by $R[M]$, and sometimes by RM, the closure of M in R, by $J_R(M)$ (sometimes J_RM, $J(M)$ or even JM) the set of all interior points of M with respect to R:

$$J_R(M) = R \setminus R(R \setminus M).$$

A closed set in R is said to be *canonical* if it is the closure of some open set; an open set is said to be *canonical* if it is the set of all interior points of a closed set. Canonical closed and canonical open sets will be denoted by A and U respectively. If $A = R(O)$, where O is open in R, and U is the set of all interior points of A, then *a fortiori*

$$A = R(U).$$

If $U = J(E)$, where E is closed in R, and A is the closure of U, then

$$U = J_R(A).$$

Thus
$$\begin{Bmatrix} U = J(A) \\ A = R(U) \end{Bmatrix} \quad \text{whenever} \quad \begin{cases} A = R(U), \\ U = J(A). \end{cases}$$

Finally, it is easy to see that canonical open (closed) sets may be defined as complements to the canonical closed (open) sets.

A closed (open) covering of the space R is called *canonical* if the elements of the covering are canonical. An *exact covering* is by definition a canonical closed covering, no two elements of which have interior points in common.

LEMMA 2·1. Any closed covering

$$\epsilon = \{E_1, \ldots, E_s\}$$

of R has an exact refinement with a nerve which is a subcomplex of the nerve of ϵ.

Proof. Let
$$O_i = R \setminus \bigcup_{k \neq i} E_k.$$

Obviously $O_i \subseteq J(E_i)$. Define $\quad A_1 = R(O_1),$

and suppose that in the closed covering

$$\epsilon_k = \{A_1, \ldots, A_k, E_{k+1}, \ldots, E_s\}$$

the elements A_i are canonical, $A_i \subseteq E_i$ and that no A_i has common interior points with any other element of ϵ_k. Define

$$\begin{aligned} O_{k+1} &= R \setminus (A_1 \cup \ldots \cup A_k \cup E_{k+2} \cup \ldots \cup E_s) \subseteq J(E_{k+1}), \\ A_{k+1} &= R(O_{k+1}), \\ \epsilon_{k+1} &= \{A_1, \ldots, A_{k+1}, E_{k+2}, \ldots, E_s\}. \end{aligned}$$

The covering ϵ_s is exact and a refinement of ϵ. Since any open covering ω possesses a closed refinement with a nerve which is a subcomplex of the nerve of ω there follows:

COROLLARY 1. Any open covering ω of R has an exact refinement with a nerve which is a subcomplex of the nerve of ω.

Furthermore

LEMMA 2·2. Any open covering ω has a canonical open refinement with a nerve which is a subcomplex of the nerve of ω.

Proof. Let $\omega = \{O_1, \ldots, O_s\}$. Take the closed $e_i \subseteq o_i$ so as to get a covering of R; then choose the open o_i under the condition

$$e_i \subseteq o_i \subseteq R(o_i) \subseteq O_i,$$

and put $U_i = JR(o_i)$. The covering

$$\{U_1, \ldots, U_s\}$$

has the required properties.

If the covering β is a refinement of the covering α but α is not a refinement of β, then β is said to follow α in the partially ordered set of all coverings of the space R (which is supposed to contain an infinite number of points). In this way the sets of all open, open canonical, closed, closed canonical, and exact coverings of R become directed sets.

We now associate with each covering α its nerve C_α, and if the covering β follows α, we put in correspondence with each element $e_{\beta i}$ of β an element $e_{\alpha i}$ of α containing $e_{\beta j}$ as a subset. This realizes a simplicial mapping ϖ_α^β of C_β in C_α, and the mappings so obtained are called projections.

In this way we obtain:

$$\left.\begin{array}{ll}
(1) & \text{The spectrum of all open}\\
(2) & \text{The spectrum of all open canonical}\\
(3) & \text{The spectrum of all closed}\\
(4) & \text{The spectrum of all closed canonical}\\
(5) & \text{The spectrum of all exact}
\end{array}\right\}\ \text{coverings of } R.$$

The spectrum of the exact coverings of R is exact and is called *the spectrum of R*.

Among the five spectra just mentioned (2) is a confinal part of (1); and (4), (5) are confinal parts of (3). It follows that (1), (2) on the one hand and (3), (4), (5) on the other have the same Δ- and ∇-groups. Furthermore, it follows from the normality of R that (1) and (3) have the same Δ- and ∇-groups (see e.g. Lefschetz 1942, chapter VII). Therefore:

All of the spectra (1)–(5) have the same Δ- and the same ∇-groups; these are called the Δ- and ∇-groups of the space R.

Let E be a closed set in R. For any covering ϵ of R the covering $E\epsilon$ of E is defined as the set whose elements are the non-empty intersections of E with the elements of ϵ. The nerve of $E\epsilon$ is a subcomplex of the nerve C_ϵ of ϵ, and this subcomplex is called EC_ϵ. The subcomplexes EC_ϵ form a subspectrum EC of the spectrum C. This is of course true if we take for the spectrum C any of the spectra (1)–(5). In particular, the spectrum EC is exact if C is the spectrum (5). In all of these cases the Δ- and ∇-groups of C mod EC are the same; they are called the Δ- and ∇-groups of R mod E.

Let f be a mapping of the space R into the space R^*. This mapping determines a homomorphism H_f^* (or H^*) of $\nabla_{R^*}^r$ in ∇_R^r and a homomorphism H_*^f (or H_*) of Δ_R^r in $\Delta_{R^*}^r$, these homomorphisms being conjugate:

$$(H_* x . x^*) = (x . H^* x^*) \quad \text{for} \quad x^* \in \nabla_{R^*}^r \quad (x \in \Delta_R^r) \tag{1}$$

(the scalar product is taken in the sense of the Pontrjagin character theory; see e.g. Alexandroff 1943, § 1).

The homomorphisms H^*, H_* can be defined on the basis of any one of the five spectra (1)–(5) and this without influence on the final result. Let us use, for instance, the spectrum of the open coverings. Take any ∇-class $z^* \in \nabla_{R^*}^r$ and any $z_{\alpha^*} \in z^*$. Let

$$\alpha^* = \{O_1^*, \ldots, O_{S^*}^*\}, \quad O_i = f^{-1} O_i^*, \quad \alpha = \{O_1, \ldots, O_S\}$$

(among the inverse images $O_i = f^{-1} O_i^*$ of course only those are taken which are not empty). The correspondence $O_i \rightarrow O_i^*$ realizes a simplicial mapping h of the nerve C_α of α into the nerve C_{α^*} of α^*, and this mapping transforms the ∇-cycle z_{α^*} into the ∇-cycle z_α defined by

$$(z_\alpha . t_\alpha^r) = (z_{\alpha^*} . h t_\alpha^r)$$

for any oriented simplex t_α^r of C_α. The ∇-class containing z_α is by definition the element $H^* z^*$ of ∇_R.

In what follows the homomorphism H_* will not be needed; as a matter of fact it is completely determined by (1).

Two important special cases of the homomorphism H^*

(1) Let R be a closed set in the space R^* and f the identical mapping of R into R^*. The corresponding homomorphisms H_f^* and H_*^f are nothing but the intersection- and the extension-homomorphisms $J_R^{R^*}$ and $E_{R^*}^R$ (operating respectively on $\nabla_{R^*}^r$ and Δ_R^r) of my previous paper (Alexandroff 1943).

An element z_R^r of ∇_R^r is called *extendable* (over R^*) if it is the image under the homomorphism $J_R^{R^*}$ of some element $z_{R^*}^r$ of $\nabla_{R^*}^r$; each of these $z_{R^*}^r$ is called an extension of the extendable element z_R^r:

$$z_R^r = J_R^{R^*} z_{R^*}^r.$$

(2) Let f be a mapping of the space R into the n-dimensional sphere S^n. As the group $\nabla_{S^n}^n$ (with the group of all integers as coefficient-group) is infinite cyclic, a generator of this group may be chosen (one of the two possible); this generator z^n is called an orientation of the sphere S^n, thus f becomes a mapping of R into the oriented sphere S^n. The image $d^n \in \nabla_R^n$ of z^n under the corresponding homomorphism H_f^* is called the degree of f. In § 5 the following two theorems will be proved:

THEOREM 2·1 (the Hopf extension theorem). Let R be a bicompactum of dimension $\leqslant n$ and f_0 a mapping of a closed set $E \subset R$ into the oriented n-sphere S^n. In order that the mapping f_0 be extendable over R it is necessary and sufficient that its degree d_0 be so extendable. Moreover, for any extension d^n of d_0^n there exists an extension f of f_0 with the degree d^n.

THEOREM 2·2 (the Hopf classification theorem). Let R be a bicompactum of dimension $\leqslant n$. The homotopy classes of mappings of R into the oriented n-sphere S^n are in $(1:1)$-correspondence with the elements z^n of ∇_R^n: to each element $z^n \in \Delta_R^n$ there correspond mappings of R in S^n with the degree z^n, and two mappings are homotopic to each other if and only if they have the same degree.

Our proof of these theorems will be practically identical with the proof by Hurewicz & Wallman (1941) for the metric case of the same theorems. The most elementary part of theorem 2·2 will be proved in this section, namely:

2·20. Two homotopic mappings of a bicompactum in a sphere have the same degree.

This theorem is a special case of

THEOREM 2·3. Let f be a mapping of the bicompactum R into the compactum R^*. Then for any element z^* of $\nabla_{R^*}^r$ there is a positive number δ with the property:

for every mapping g of R in R^* whose distance from f is $\leqslant \delta$

$$H_f^* z^* = H_g^* z^*.$$

Theorem 2·3 admits a further generalization which will be given after an auxiliary definition.

Suppose the spaces R, R^* are given and let ω^* be an open covering of R^*. Two mappings f, g of R into R^* are said to be ω^*-modifications of each other if for every point $x \in R$ the two points $f(x)$ and $g(x)$ belong to some element of ω^*. Now the following theorem will be proved (which implies 2·3 and *a fortiori* 2·20):

THEOREM 2·4. Let f be a mapping of the bicompactum R into the bicompactum R^*. Then for any element z^* of $\nabla_{R^*}^r$ there exists an open covering

$$\omega^* = \{o_1^*, \ldots, o_u^*\}$$

of R^* with the following property: If g is any ω^*-modification of f then $H_f^* z^* = H_g^* z^*$.

Proof. Let $z_{\alpha^*} \in z^*$ with $\alpha^* = \{O_1^*, \ldots, O_{s^*}^*\}$. Take the covering

$$\alpha_f = \{O_1^f, \ldots, O_s^f\}, \quad \text{where} \quad O_i^f = f^{-1} O_i^* \neq \Lambda.$$

There exists a covering

$$\beta = \{O_{\beta 1}, \ldots, O_{\beta s}\} \quad \text{with} \quad RO_{\beta i} \subseteq O_i^f,$$

and a covering $$\omega^* = \{o_1^*, \ldots, o_u^*\}$$

with the property: $$\text{if} \quad f(RO_{\beta i}) \cap o_j^* \neq \Lambda \quad \text{then} \quad o_j^* \subseteq O_i^*.$$

Now let g be an ω^*-modification of f. If $x \in RO_{\beta i}$ then both $f(x)$ and $g(x)$ are points of an o_j^* and $f(x) \in f(RO_{\beta i}) \cap o_j^*$; thus

$$o_j^* \subseteq O_i^*, \; g(x) \in O_i^*.$$

This implies $RO_{\beta i} \subseteq g^{-1}O_i^* = O_i^g$. Thus β is a refinement of both α_f and $\alpha_g = \{O_i^g\}$. It will be shown that $z_{\alpha *}$ is carried by both $\pi_\beta^{\alpha_f} H_f^*$ and $\pi_\beta^{\alpha_g} H_g^*$ into the same z_β. In fact, let $t_\beta^r = (O_{\beta i_0}, \ldots, O_{\beta i_r})$ be a simplex of the nerve C_β of β. Then

$$t_f^r = (O_{i_0}^f, \ldots, O_{i_r}^f) \quad \text{and} \quad t_g^r = (O_{i_0}^g, \ldots, O_{i_r}^g)$$

are simplices of C_{α_f} and C_{α_g} respectively, and the values of $\pi_\beta^{\alpha_f} H^* z_{\alpha *}$ and of $\pi_\beta^{\alpha_g} H_g^* z_{\alpha *}$ on t_β^r are equal to the corresponding values of $H_f^* z_{\alpha *}$ and $H_g^* z_{\alpha *}$ on t_f^r and t_g^r. But the latter ones are equal to the value of $z_{\alpha *}$ on $t_{\alpha *}^r = (O_{i_0}^*, \ldots, O_{i_r}^*)$. Thus the ∇-class of R containing z_β is at the same time $H_f^* z^*$ and $H_g^* z^*$, Q.E.D.

3. The bicompact case

Let $C = \{C_\alpha, \varpi_\alpha^\beta\}$ be an exact spectrum; if in each $C_\alpha \in C$ an element $t_\alpha \in C_\alpha$ is taken in such a way that $t_\alpha = \varpi_\alpha^\beta t_\beta$ whenever $\beta > \alpha$, then the system $\xi = \{t_\alpha\}$ so obtained is called a *thread* of the spectrum C; the thread $\xi' = \{t'_\alpha\}$ is called a part (or a subthread) of the thread $\xi = \{t_\alpha\}$ if $t'_\alpha \leqslant t_\alpha$ for all α. A thread ξ is called *maximal* if ξ is not a part of any thread different from it.

An exact spectrum $C = \{C_\alpha, \varpi_\alpha^\beta\}$ defines a bicompact space R_C in the following way (Kurosch 1935). The maximal threads of C are the points of R_C; the neighbourhood $O_{\alpha_0}(\xi)$ of $\xi = \{t_\alpha\}$ is defined as the set of all points $\xi' = \{t'_\alpha\}$ such that $t'_{\alpha_0} \leqslant t_{\alpha_0}$. The attention of the reader is called to the following fundamental theorem:

Theorem 3·0 (Kurosch 1935). Let R be a bicompactum and $C = \{C_\alpha, \varpi_\alpha^\beta\}$ its spectrum. To any maximal thread $\xi = \{t_\alpha\}$ of C there corresponds the point

$$f(\xi) = \bigcap_\alpha E(t_\alpha) \in R,$$

where $E(t_\alpha)$ is the intersection of the elements of α corresponding to the vertices of t_α. The function f establishes a homeomorphism between the spaces R_C and R.

Thus the maximal threads of C are in $(1:1)$-correspondence with the points of R. Now a $(1:1)$-correspondence will be constructed between the closed subsets of R and certain subspectra of C called maximal. In construction this $C = \{C_\alpha, \varpi_\alpha^\beta\}$ still denotes the spectrum of the bicompactum R.

Lemma 3·1. Let t_α belong to C_α, α being fixed. There always exists a thread containing t_α as an element.

Proof. For any $\beta > \alpha$ take all simplices belonging to the inverse image of t_α under ϖ_α^β (i.e. all t_β satisfying $\varpi_\alpha^\beta t_\beta = t_\alpha$). Now take all projections of the simplices just obtained. Denote by Ξ the resulting set of simplices. I assert that the only element of Ξ belonging to C_α is t_α. In fact, if $t'_\alpha \in \Xi$, then by definition $t'_\alpha = \varpi_\alpha^\beta t_\beta$, where either $\varpi_\alpha^\beta t_\beta = t_\alpha$ and thus $t_\alpha = t'_\alpha$, or $t_\beta = \varpi_\beta^\gamma t_\gamma$ with $\gamma > \beta$ and $\varpi_\alpha^\gamma t_\gamma = t_\alpha$. Then $\gamma > \beta > \alpha$ and $t'_\alpha = \varpi_\alpha^\beta t_\beta = \varpi_\alpha^\beta \varpi_\beta^\gamma t_\gamma = \varpi_\alpha^\gamma t_\gamma = t_\alpha$. The set Ξ has, moreover, the property that the projection of any element of Ξ still belongs to Ξ. Finally it can be shown that for any $t_\beta \in \Xi$ and $\gamma > \beta$ there exists a $t_\gamma \in \Xi$ with $\varpi_\beta^\gamma t_\gamma = t_\beta$. This is evident for $\beta = \alpha$.

Let $\beta \neq \alpha$. Then $t_\beta = \varpi^{\gamma'}_{\beta} t_{\gamma'}$, where $\varpi^{\gamma'}_{\alpha} t_{\gamma'} = t_\alpha$. Take $\gamma'' > \gamma, \gamma'$ and $t_{\gamma''}$ with $\varpi^{\gamma''}_{\gamma'} t_{\gamma''} = t_{\gamma'}$. Then $\varpi^{\gamma''}_{\alpha} t_{\gamma''} = \varpi^{\gamma'}_{\alpha} \varpi^{\gamma''}_{\gamma'} t_{\gamma''} = t_\alpha$; thus $t_{\gamma''} \in \Xi$ and $\varpi^{\gamma''}_{\gamma} t_{\gamma''} \in \Xi$. Set $\varpi^{\gamma''}_{\gamma} t_{\gamma''} = t_\gamma$. Then

$$\varpi^{\gamma}_{\beta} t_\gamma = \varpi^{\gamma}_{\beta} \varpi^{\gamma''}_{\gamma} t_{\gamma''} = \varpi^{\gamma'}_{\beta} \varpi^{\gamma''}_{\gamma'} t_{\gamma''} = \varpi^{\gamma'}_{\beta} t_{\gamma'} = t_\beta.$$

Thus Ξ satisfies the conditions of Kurosch's lemma of his paper (1935, p. 472), and therefore contains a thread ξ; as the only element of Ξ belonging to C_α is t_α, this t_α necessarily belongs to ξ, Q.E.D.

Let us call the closed subcomplex $A'_\alpha \subseteq C_\alpha$ an *extension of the subcomplex* A_α of C_α if $A_\alpha \subseteq A'_\alpha$, and each maximal simplex of A'_α belongs to the open star of some simplex of A_α. The subcomplex A itself is by definition its own trivial extension.

The subspectrum $A' = \{A'_\alpha, \varpi^{\beta}_{\alpha}\}$ of the spectrum $C = \{C_\alpha, \varpi^{\beta}_{\alpha}\}$ is called an extension of the subspectrum $A = \{A_\alpha, \varpi^{\beta}_{\alpha}\}$ of C if every A'_α is an extension of A_α (as subcomplex of C_α). Again, the subspectrum A is its own trivial extension. The subspectrum A of C is called *maximal* if it has no extensions except the trivial one.

The purpose of the present section is to prove the following three assertions:

(1) Any maximal subspectrum $A = \{A_\alpha, \varpi^{\beta}_{\alpha}\}$ of the spectrum $C = \{C_\alpha, \varpi^{\beta}_{\alpha}\}$ satisfies the condition:—

($\mathscr{S}$). Every thread of A which is maximal in A is maximal in C.

(2) Every subspectrum A of C satisfying the condition ($\mathscr{S}$) can be put in the form

$$A = EC,$$

where E is a closed subset of the bicompactum R.

(3) Every subspectrum A of C of form $A = EC$ (with a closed $E \subseteq R$) is maximal.

These three assertions are obviously equivalent to the following two theorems:

THEOREM 3·1. The condition ($\mathscr{S}$) is necessary and sufficient in order that the subspectrum A be maximal.

THEOREM 3·2. In order that a subspectrum A of the spectrum C be of the form $A = EC$ it is necessary and sufficient that A be maximal.

Before proceeding to the proof of these theorems we note that to different sets E there correspond different subspectra EC. In fact, if x is a point which belongs, say, to E' but not to E'', let an (exact) covering α be chosen with an element e_α containing the point x but no point of E''. Then the vertex e_α belongs to $C_{E'\alpha} = A'_\alpha$ and does not belong to $C_{E''\alpha} = A''_\alpha$. Thus the subspectra $A' = E'C$ and $A'' = E''C$ are different. Therefore theorem 3·2 shows that the operator EC (with variable E) realizes a (1 : 1)-correspondence between the system of the closed subsets of R and that of the maximal subspectra of C.

The proof of (1) is rather obvious: let $x = \{t_\alpha\}$ be a thread which is maximal in A but not in C. Let $x' = \{t'_\alpha\}$ be a non-trivial extension of this thread; add to each A_α the simplex t'_α with all its faces. The complexes A'_α so obtained form a non-trivial extension of A.

Proof of (2). Let $A = \{A_\alpha, \varpi^{\beta}_{\alpha}\}$ be a subspectrum of the spectrum $C = \{C_\alpha, \varpi^{\beta}_{\alpha}\}$. Suppose that A satisfies the condition ($\mathscr{S}$). Then to each maximal thread of the spectrum A there corresponds a point of R; the point set E of R so obtained, being obviously homeomorphic to the space defined by A, is thus bicompact and therefore closed in R. It remains to show that $A = EC$. To this end take any

$$t_\alpha = (e_{\alpha 0} \dots e_{\alpha r}) \in EC_\alpha,$$

and a point $x \in E \cap e_{\alpha 0} \cap \dots \cap e_{\alpha r}$. By the definition of E the point x may be identified with a maximal thread $\xi = (\dots t'_\alpha \dots)$ of A; the simplex $t'_\alpha \geqslant t_\alpha$ is a simplex of A_α; hence $t_\alpha \in A_\alpha$. Since this is proved for an arbitrary $t_\alpha \in EC_\alpha$, $EC_\alpha \subseteq A$. To prove the inclusion $A \subseteq EC_\alpha$, consider any $t_\alpha \in A_\alpha$ and take a maximal thread

$$x = (\dots t'_\alpha \dots) \text{ of } A,$$

with $t'_\alpha \geqslant t_\alpha$ (such a thread exists by lemma 3·1). The point $x \in E$ is contained in all elements of α corresponding to the vertices of t'_α, in particular, to the vertices of t_α; therefore $t_\alpha \in EC_\alpha$, Q.E.D.

Proof of (3). Suppose that the subspectrum $A = EC = \{A_\alpha, \varpi_\alpha^\beta\}$ is not maximal. Let $A' = \{A'_\alpha, \varpi_\alpha^\beta\}$ be a non-trivial extension of A; in particular, let for a fixed α the subcomplex A'_α be a non-trivial extension of A_α. Finally, let

$$t_\alpha = (e_{\alpha 0} \dots e_{\alpha r}) \in A_\alpha,$$
$$t'_\alpha = (e_{\alpha 0} \dots e_{\alpha r}\, e_{\alpha, r+1}) \in A'_\alpha \setminus A_\alpha.$$

The vertices $e_{\alpha 0}, \dots, e_{\alpha r}$ correspond both to the elements $E_{\alpha 0}, \dots, E_{\alpha r}$ of α and to the elements $E \cap E_{\alpha 0}, \dots, E \cap E_{\alpha r}$ of $E\alpha$. According to lemma 3·1 take a thread

$$\xi' = (\dots t'_\alpha \dots)$$

of the spectrum A'. Let, for any $\beta > \alpha$,

$$t'_\beta = (e_{\beta 0} \dots e_{\beta r} \dots e_{\beta u_\beta}) \in \xi'.$$

The spectrum A' being an extension of A, there is certainly a simplex t''_β in C_β satisfying both

$$t''_\beta \geqslant t'_\beta \quad \text{and} \quad t''_\beta \geqslant t_\beta \in A_\beta.$$

In other words, any element of β corresponding to a vertex of t'_β (even to a vertex of t''_β) has common points with an element of β intersecting E. Now take a maximal extension $\xi'' = (\dots t''_\alpha \dots t''_\beta \dots)$ (in C) of ξ'. The thread ξ'' may be identified with a point x of R which for any $\beta > \alpha$ belongs to an element of β having common points with some element intersecting E. As this is true for every $\beta > \alpha$, it is seen that $x \in E$. On the other hand, x belongs to all elements of α corresponding to vertices of t''_α, and, in particular, to those of t'_α. Hence it follows that $t'_\alpha \in A_\alpha$, contrary to our hypothesis. Thus the assertion (3), and with it theorems 3·1 and 3·2, are completely proved.

Remark. Since the spectrum C (being the exact spectrum of the bicompactum R) is a Hausdorff spectrum, any thread of C is contained in only one maximal thread. Hence

COROLLARY 3·1. If A is a maximal subspectrum of C, then A contains, together with any thread, its maximal extension (in C).

Consider in the spectrum $C = \{C_\alpha, \varpi_\alpha^\beta\}$ a system of closed subcomplexes $A_\alpha \in C_\alpha$ (for all α) such that for every $\beta > \alpha$ the projection ϖ_α^β maps A_β *into* A_α (not necessarily *on* to A_α). Such systems (with the corresponding projections) will be called for a moment 'incomplete' subspectra of C. In order that an incomplete subspectrum $A = \{A_\alpha, \varpi_\alpha^\beta\}$ be a subspectrum it is necessary and sufficient that every $t_\alpha \in A_\alpha \in A$ can be included in a thread $\xi = (\ldots t_\alpha \ldots)$ (of A). Two incomplete subspectra are called equivalent if they have the same threads and the incomplete subspectrum $A' = \{A'_\alpha, \varpi_\alpha^\beta\}$ is called a reduction of $A = \{A_\alpha, \varpi_\alpha^\beta\}$ if A and A' are equivalent and $A'_\alpha \subseteq A_\alpha$ for all α. Applying the reduction process transfinitely for the incomplete subspectrum $A = \{A_\alpha, \varpi_\alpha^\beta\}$ a reduction $A^\tau = \{A_\alpha^\tau, \varpi_\alpha^\beta\}$ is obtained which is a subspectrum of C. It is easily seen that each A_α^τ has as elements all those and only those $t_\alpha \in A_\alpha$ which are contained in the threads $\xi = (\ldots t_\alpha \ldots)$ of A. Thus, the *complete reduction* A^τ of the incomplete subspectrum A is uniquely determined.

Now let two subspectra $A' = \{A'_\alpha, \varpi_\alpha^\beta\}$ and $A'' = \{A''_\alpha, \varpi_\alpha^\beta\}$ of the spectrum C be given. Then the complexes $A'''_\alpha = A'_\alpha \cap A''_\alpha$ (with the projections ϖ_α^β) form an incomplete subspectrum $A''' = \{A'''_\alpha, \varpi_\alpha^\beta\}$. Its complete reduction is called the intersection of the subspectra A' and A'' and is denoted by $A' \cap A''$.

Now, from corollary 3·1 there follows

COROLLARY 3·2. The intersection of two maximal subspectra of C is a maximal subspectrum.

Moreover:

COROLLARY 3·3. If $A' = E'C$, $A'' = E''C$, and $A = A' \cap A''$, $E = E' \cap E''$, then $A = EC$.

Proof of Corollaries 3·2 *and* 3·3. Any maximal thread of A is a thread of both A' and A''; these subspectra being maximal, the maximal extension of ξ (with respect to C) is contained in A' as well as in A'' and thus in A; therefore ξ is a maximal thread in C, which proves the maximality of the subspectrum A. Furthermore, as has just been shown, the point $x \in R$ corresponding to the given maximal thread ξ of A belongs to both E' and E''. On the other hand, if $x \in E$ then the maximal thread corresponding to x is a thread of both A' and A'', and thus a thread of A. In other words, the maximal threads of A are just the threads corresponding to the points of E. This, together with the argument used in the proof of the assertion (3), completes the proof of corollary 3·3. Now we say that the spectrum $C = \{C_\alpha, \varpi_\alpha^\beta\}$ is the sum of its maximal subspectra $C' = \{C'_\alpha, \varpi_\alpha^\beta\}$ and $C'' = \{C''_\alpha, \varpi_\alpha^\beta\}$ if $C_\alpha = C'_\alpha \cup C''_\alpha$ for each α.

With these definitions, *the* (1:1)-*correspondence between the maximal subspectra of the spectrum C (of the bicompactum R) and the closed subsets of R is an isomorphism with respect to both addition and intersection.*

4. DIMENSION

The dimension of the nerve of a covering is called the dimension of the covering.

Let R be a normal space. If each open covering ω of R has an n-dimensional (open) refinement ϵ, R is said to be at most n-dimensional. In this definition we may suppose ϵ to be closed instead of open, or to be open canonical, closed canonical, exact. If R is at most n-dimensional but not at most $(n-1)$-dimensional, the dimension of R is said to be n:

$$\dim R = n.$$

It is one of the main purposes of this paper to prove

THEOREM 4·1. If the dimension of the normal space is n, then the exact spectrum C of R contains a maximal subspectrum C_0 such that†

$$\nabla^n(C \bmod C_0, J) \neq 0 \quad (\text{dually: } \Delta^n(C \bmod C_0, \bar{J}) \neq 0),$$

while for $r > n$, for any maximal subspectrum C_0 and for any coefficient-group, the r-dimensional Betti groups of $C \bmod C_0$ are zero.

THEOREM 4·2. If the dimension of the normal space R is n, then R contains a closed (and even closed canonical) set E such that

$$\nabla^n(R \bmod E, J) \neq 0 \quad (\text{dually: } \Delta^n(R \bmod E, \bar{J}) \neq 0),$$

while for $r > n$, for any closed $E \subseteq R$ and any coefficient-group the r-dimensional Betti groups of $R \bmod E$ are zero.

Owing to theorem 3·2 the theorems 4·1 and 4·2 are immediately equivalent if R is bicompact. In this special case both theorems will be proved in § 5; the general case will be solved in § 6. In the present section some important but more elementary theorems will be proved. In the proofs given I often follow my paper (1940).

Preliminary definition. A closed E in the normal R is called *correct* if there exists a continuous real-valued function f on R such that $f(x) = 0$ if and only if $x \in E$. One can easily see that the correct closed sets are nothing but those closed sets which are intersections of an enumerable system of open sets. The complements $R \setminus E$ of correct closed sets $E \subseteq R$ are called correct open sets.

LEMMA 4·1. The closed set E and its neighbourhood $O(E)$ are arbitrarily chosen in the normal space R. Then there exists a correct closed set A which contains E and is contained in $O(E)$.

Proof. By the classical Urysohn lemma there exists a continuous function f defined on the whole space R and such that

$$\begin{cases} f(x) = 0 & \text{on } E, \\ f(x) = 1 & \text{on } R \setminus O(E). \end{cases}$$

Define A as the set of all $x \in R$ with $f(x) = 0$.

---

† The additive group of all integers is denoted by J, its character-group (i.e. the continuous group of all real numbers reduced mod. 1) is denoted by $\bar{J}$.

It will be convenient for us to bring lemma 4·1 into the following form:

LEMMA 4·2. For any open set O in the normal space R and for an arbitrary closed $E \subseteq O$ there exists a correct open set U such that $E \subseteq U \subseteq O$.

Proof. Take the correct closed set A under the condition $R \setminus O \subseteq A \subseteq R \setminus E$, then $U = R \setminus A$ is a correct, open set, and $E \subseteq U \subseteq O$.

An open (closed) covering whose elements are correct is itself called correct. From lemmas 4·2 and 4·1 follow respectively:

LEMMA 4·3. To any open (closed) covering

$$\epsilon = \{M_1, \ldots, M_s\}$$

there corresponds a correct open (closed) covering

$$\epsilon' = \{M'_1, \ldots, M'_s\},$$

such that in the open case $M_i \supseteq M'_i$, while in the closed case $M_i \subseteq M'_i$; and the coverings ϵ and ϵ' have the same nerve. More precisely: if the closed set $E_i \subseteq M_i$ (the open set $O_i \supseteq M_i$) are given for $i = 1, 2, \ldots, s$, then the M'_i can be chosen to satisfy the conditions $E_i \subseteq M'_i \subseteq M_i$ in the open and $M_i \subseteq M'_i \subseteq O_i$ in the closed case.

Let ω be an open covering of the normal space R and f a continuous mapping of R in a space R'. This mapping is called an ω-mapping if every point $x' \in R'$ has a neighbourhood whose inverse image under f is contained in some element of ω.

Remark. In the case of a closed mapping f it is sufficient to suppose that the inverse image of every point $x' \in R'$ is contained in some $o \in \omega$. In fact, for $f^{-1}(x') \subseteq o \in \omega$ the set $f(R \setminus o)$ is closed in R' and $R' \setminus f(R \setminus o)$ is a neighbourhood of x' with an inverse image lying in o.

Let $R' = \bar{C}$ be a polytope given in a definite simplicial decomposition C; among the ω-mappings of the normal space R into R' of special interest to us are those which satisfy the condition:

The elements o_i of $\omega = \{o_1, \ldots, o_s\}$ are in (1 : 1)-correspondence with the vertices e_i of C, and each o_i is the inverse image of the open star of the corresponding vertex e_i of C (it follows that C is the nerve of ω).

These ω-mappings will be called (ω, C)-*mappings*;[†] they are characterized by the following construction:

Suppose that for $i = 1, \ldots, s$ the continuous non-negative functions $\mu_i(x)$ are defined on R by the condition that $\mu_i(x) = 0$ if and only if $x \in R \setminus o_i$ (this implies that ω is correct). Define for any $x \in R$ the point $f(x)$ as the centre of the masses $\mu_i(x)$ placed respectively at the points e_i. This centre lies in the simplex of C spanned by the vertices e_i corresponding to the elements of ω containing x. The mapping f so defined is an (ω, C)-mapping; conversely, every (ω, C)-mapping can be obtained in this way: to get the $\mu_i(x)$ take the (open) simplex $(e_{i_0} \ldots e_{i_r})$ to which $f(x)$ belongs and set $\mu_i(x) = 0$ if i differs from all $i_0, \ldots, i_r$, while for $i = i_k$ the value $\mu_i(x)$ is defined as the barycentric coordinate of $f(x)$ corresponding to the vertex e_{i_k} in $(e_{i_0} \ldots e_{i_r})$. This same construction leads to

---

[†] After Kuratowski I studied these mappings in 1940; Hurewicz & Wallman call them 'barycentric' ω-mappings.

Lemma 4·4. Let ω be a correct open covering of the normal space R; let C be the nerve of ω and $R' = \bar{C}$ a polytope with the simplicial decomposition C. Then there exists an (ω, C)-mapping of R into R'.

This result, together with lemma 4·3, implies

Lemma 4·5. To any open covering ω of an n-dimensional normal space correspond ω-mappings into n-dimensional polytopes.

Using the Wallman-Hurewicz procedure of their p. 73 the mappings *into* in Lemma 4·5 may be replaced by mappings *on* to.

On the other hand:

Lemma 4·6. For every n-dimensional R there exists such an open covering ω that no ω-mapping of R into a bicompactum of dimension $\leqslant n-1$ is possible.

Proof. Suppose that the contrary is true and that to any open covering ω of the given R there corresponds a bicompactum R'_ω of dimension $\leqslant n-1$ and an ω-mapping f_ω of R into R'_ω. Take for any point $x' \in R'_\omega$ a neighbourhood $o_{x'}$ of x' such that its inverse image $f^{-1}o_{x'}$ lies in some element of ω. Choose a finite covering $\{o'_1, \ldots, o'_{s'}\} = \omega'$ of R' composed of some of the above $o_{x'}$. Let $\omega'' = \{o''_1, \ldots, o''_{s''}\}$ be an $(n-1)$-dimensional refinement of ω'. The inverse images $o_i = f^{-1}o''_i$ form an $(n-1)$-dimensional refinement of ω, contradicting the definition of ω.

The following fundamental theorem has just been proved:

Theorem 4·3. The dimension of the normal space R is the least among the integers n with the property: for an arbitrary open covering ω of R there exists an ω-mapping of R on to an n-dimensional polytope. In this formulation the words 'on to' may be replaced by 'into'.

Let f be a mapping of the normal space R into R', where R' is the n-dimensional Euclidean or spherical space, or Hilbert space. Suppose that there is given a positive number ϵ and a correct open covering ω of R which is 'fine enough', in the sense that $\rho(f(x), f(y)) < \epsilon$ whenever x and y belong to the same element of ω (Hurewicz & Wallman 1941, p. 63). Choose in each element $o_i \in \omega$ a point p_i and take the vertices e_i of the nerve C at or near the points $f(p_i)$. Define the functions $\mu_i(x)$ as above, and denote by $f_\epsilon(x)$ the centre of the masses $\mu_i(x)$ placed in e_i. In this way a continuous mapping f_ϵ of R into the polytope $\bar{C}$ is obtained (which is a geometric realization of the nerve C in the space R'). It is easily seen that for an arbitrary $x \in R$

$$\rho(f(x), f_\epsilon(x)) < \epsilon. \tag{1}$$

Thus

Approximation Theorem 4·4. To any $\epsilon > 0$ and to any continuous mapping of the r-dimensional normal space R into the spherical or Euclidean or Hilbert space R' there exists a continuous mapping f_ϵ of R into an r-dimensional polytope imbedded in R', the mapping f_ϵ being an ϵ-approximation of f in the sense that (1) holds for all $x \in R$. The polytope $\bar{C}$ may be chosen as the geometric realization of the nerve C of any sufficiently fine correct open covering of R.

From this approximation theorem there follows immediately:

LEMMA 4·7. For $r < n$ any continuous mapping of an r-dimensional normal space into the n-dimensional sphere S^n is inessential; thus, any two such mappings are homotopic to each other.

Furthermore, we shall prove by the methods of Alexandroff & Hopf (1935, pp. 374–375):

LEMMA 4·8. If f is an essential mapping of the normal space R on the closed simplex T, then there exists an $\epsilon > 0$ such that any mapping of R into T differing from f by less than ϵ covers a fixed simplex concentric with T. Thus for $r < n$ there is no essential mapping of an r-dimensional space on an n-dimensional closed simplex.

Proof of Lemma 4·8. Let f be an essential mapping of the normal R on the closed spherical region E^n of radius 1 and centre o. Denote by e^n, E'^n the spherical regions of radii 1/3, 2/3, respectively, and the same centre o; the boundaries of E^n, E'^n, e^n are the spheres S, S', s. Denote finally by H the closed region between the spheres S and S', and set

$$\Gamma = f^{-1}(H).$$

Let f_1 be any mapping of R in E^n differing from f by less than 1/3. Then for $x \in \Gamma$ the rectilinear segment $\overline{f(x)f_1(x)}$ has no points in common with e^n.

Now define the mapping f' as follows. For $x \in \Gamma$ draw the line $\overline{of(x)}$ and take the portion δ of this line lying between S and S'; the segment δ is considered as directed from S to S' (i.e. towards the centre o). Now $f'(x)$ is defined as the point which divides the segment $\overline{f(x)f_1(x)}$ in the same ratio as $f(x)$ divides the segment δ.

For $x \in R \setminus \Gamma$ one simply sets $f'(x) = f_1(x)$. This definition implies that for $x \in f^{-1}(S)$.

$$f'(x) = f(x),$$

and hence, since f is essential, $\quad f'(R) = E^n \supset e^n.$

But if $f'(x) \in e^n$, then certainly $x \in R \setminus \Gamma$, and therefore $f_1(x) = f'(x)$. Consequently $e^n \subseteq f_1(R)$, Q.E.D.

We now return to lemma 4·4 and consider an (ω, C)-mapping of R in the nerve C of ω. With Hurewicz & Wallman (1941, p. 73) we find a mapping f of R on to a certain $\bar{C}'$, $C' \subseteq C$ which is quasi-barycentric (i.e. the inverse image of any open star of C' is contained in some element of ω) and minimal in the sense that there is no quasi-barycentric mapping of R into a C'' with $C'' \subset C'$. This implies (by virtue of my 'sweeping out' reasoning (Alexandroff & Hopf 1935, p. 365)) that every n-dimensional simplex of C' is essentially covered by the image of R under f. Hence

THEOREM 4·5. The dimension of a normal space R is, if finite, the greatest among those integers n which allow an essential mapping of R on to an n-dimensional closed simplex.

Remark. There are sets M lying in a normal (and even bicompact) R and having a dimension greater than that of R; if M is closed in R, then necessarily $\dim M \leqslant \dim R$. A closed M in a normal R being always normal, we are led to the following problem:

Problem. Let $M \subset R$ be normal (if taken with the 'natural' topology induced by the topology of R). Does then the inequality $\dim M \leqslant \dim R$ necessarily hold?

In any case the following definition will be useful to us:

DEFINITION. Let M be a subset of the normal R. The greatest among all numbers $\dim E$, where $E \subseteq M$ is closed in R, is called the relative dimension of M with respect to R and is denoted by $\mathrm{rd}_R M$.

Remark. If R is bicompact then $\mathrm{rd}_R M$ is equal to the greatest among all $\dim E$, where E is *any bicompactum lying in M*. Thus in this case $\mathrm{rd}_R M$ is an absolute invariant of M and may be denoted simply by $\mathrm{rd}\, M$.

With the notion of the relative dimension a theorem of Hurewicz & Wallman (1941, their corollary on p. 84) may be generalized as follows:

THEOREM 4·6. *Let E be closed in the normal R. If $\mathrm{rd}_R(R \setminus E) \leqslant n$, then every mapping of E into the n-dimensional sphere S^n can be extended over R.*

Proof. Let f be the given mapping of E into S^n. By the Brouwer-Urysohn extension theorem, f can be extended over a neighbourhood U of E to a certain mapping f'. Take the neighbourhoods U_1 and U_2 of E satisfying the conditions

$$E \subseteq U_2 \subseteq RU_2 \subseteq U_1 \subseteq RU_1 \subseteq U,$$

and consider the function f' on $(RU_1) \cap (R \setminus U_2)$. The latter set is closed in $R \setminus U_2$, which is closed in R and therefore at most n-dimensional. Thus, by theorem 4·5,[†] the mapping f' of $(RU_1) \cap (R \setminus U_2)$ into S^n is extendable to a mapping f'' of $R \setminus U_2$. Defining

$$f'''(x) = f(x) \quad \text{on} \quad RU_2,$$
$$f'''(x) = f''(x) \quad \text{on} \quad R \setminus U_2,$$

the desired extension of f is found.

The theorem just proved will play an important role in the last section of this paper.

To finish the present section the following proposition will be proved, which is important for many deformation-theorems and in particular for the Hopf classification theorem as well as for the theorem on Cantor manifolds:

THEOREM 4·7. *The topological product $[R, I]$ of an n-dimensional bicompactum R and the rectilinear segment $I = <0; 1> = \{0 \leqslant t \leqslant 1\}$ is of dimension $\leqslant n+1$.*

Before proceeding to the proof of this theorem I shall call the attention of the reader to two questions which naturally arise in connexion with it:

(1) Is it possible to strengthen the inequality $\dim [R, I] \leqslant n+1$ to the equality $\dim [R, I] = n+1$?

[†] Here theorem 4·5 is used in its following form which is immediately equivalent to the previous one:

THEOREM 4·50. $\dim R \leqslant n$ *holds if, and only if, every mapping of every closed $E \subset R$ into S^n can be extended over R.*

(2) Does theorem 4·7 remain true for an arbitrary normal space R?

Proof of Theorem 4·7. Take an open covering $\Omega = \{O_h\}$ of the bicompactum $[R, I]$. It will be proved that Ω has a closed refinement of dimension $\leqslant n+1$.

To this end, choose for any point $(x,t) \in [R, I]$ an element $O_{i(x,t)}$ of Ω containing (x,t), and take an open set $o_{x,t} \subseteq R$ and an interval $d_{x,t} \subseteq I$ in such a way that

$$(x,t) \in [o_{x,t}, d_{x,t}] \subseteq O_{i(x,t)}.$$

The point x being considered as fixed, choose a finite number of such products $[o_{x,t}, d_{x,t}]$, so as to get a covering of the rectilinear segment $[x, I] \subseteq [R, I]$. Let these $[o_{x,t}, d_{x,t}]$ be

$$[o_{x,t_1}, d_{x,t_1}], \quad \ldots, \quad [o_{x,t_r}, d_{x,t_r}].$$

Set $$o_{x,t_1} \cap \ldots \cap o_{x,t_r} = o_x.$$

Then for any t, $0 \leqslant t \leqslant 1$, $[o_x, t]$ is contained in some element of Ω.

Now, considering $x \in R$ as variable, choose a finite number of the o_x just defined so as to obtain a covering

$$\omega = \{o_1, \ldots, o_q\}$$

of R. This covering still has the property that for any o_i and any t, $0 \leqslant t \leqslant 1$, $[o_i, t]$ is contained in some element of Ω. Let

$$\epsilon = \{e_1, \ldots, e_s\}$$

be a closed n-dimensional refinement of ω. As before, each of the $[e_i, t]$ lies in some $o_i \in \Omega$. A positive δ can be found in such a way that for any interval $\eta \subset I$ of length $\leqslant \delta$ and an arbitrary $e_i \in \epsilon$ the point set $[e_i, \eta]$ in $[R, I]$ is contained in some $o_i \in \Omega$. Now denote by

$$\sigma = \{0, b_1, \ldots, b_m, 1\},$$

a subdivision of the segment $I = <0, 1>$ into consecutive segments

$$\langle 0, b_1 \rangle, \quad \langle b_1, b_2 \rangle, \quad \ldots, \quad \langle b_m, 1 \rangle$$

of length $< \delta$. Take s such subdivisions

$$\sigma_i = \{0, b_1^i, \ldots, b_m^i, 1\}$$

under the sole condition that for $i \neq j$ and arbitrary h, k, b_h^i and b_k^j are always unequal. Denote by η_0^i, η_k^i (with $k = 1, \ldots, m-1$) the segments $\langle 0, b_1^i \rangle$, $\langle b_k^i, b_{k+1}^i \rangle$ and $\langle b_m^i, 1 \rangle$ respectively and set

$$E_{ik} = [e_i, \eta_k^i], \quad i = 1, 2, \ldots, s; \quad k = 0, 1, \ldots, n-1.$$

These E_{ik} form a closed covering E of $[R, I]$ which is a refinement of Ω. It will now be proved that the dimension of the covering E is $\leqslant n+1$. To do this, first note that

(1) $E_{ih} \cap E_{ik} \neq \Lambda$ only in the case $|h-k| = 1$.

(2) If $(x,t) \in E_{ih} \cap E_{i,h+1}$, then $t = b_{i,h+1}$.

(3) If $i_1, \ldots, i_{n+2}$ are all distinct, then for an arbitrary choice of $h_1, \ldots, h_{n+2}$

$$E_{i_1 h_1} \cap \ldots \cap E_{i_{n+2} h_{n+2}} = \Lambda.$$

Now let $(x, t) \in E_{i_1 h_1} \cap \ldots \cap E_{i_{n+3}\, h_{n+3}}$. Then at least two among the i's, say i_1, i_2, are equal, $i_1 = i_2 = i$ and $t = b^i_{h+1}$, where $h = h_1$, $h_2 = h + 1$, and

$$(x, t) = (x, b^i_{h+1}) \in E_{ih} \cap E_{i, h+1} \cap E_{i_3 h_3} \cap \ldots \cap E_{i_{n+3}\, h_{n+3}}.$$

But b^i_{h+1} is different from all $b^{i_3}_k, \ldots, b^{i_{n+3}}_k$, and therefore the numbers $i, i_3, \ldots, i_{n+3}$ are all distinct. Thus the $n + 2$ point-sets

$$E_{ih}, E_{i_3 h_3}, \ldots, E_{i_{n+3} h_{n+3}}$$

have a non-empty intersection, in contradiction to assertion (3). This completes the proof of theorem 4·7.

COROLLARY. Let R be a bicompactum and M a point-set in R with $\operatorname{rd} M \leqslant m$. Then, for $R' = [R, I]$ and $M' = [M, I] \subseteq R'$

$$\operatorname{rd} M' \leqslant m + 1.$$

Proof. Let $E' \subseteq M'$ be closed in R' (and therefore a bicompactum). Denote by E the projection of E' in R. Then, since E' is closed in $[E, I]$,

$$\dim E' \leqslant \dim [E, I] \leqslant m + 1, \quad \text{Q.E.D.}$$

5. PROOF OF THEOREMS 2·1, 2·2, 4·2 (FOR BICOMPACTA)

First some preliminary remarks.

LEMMA 5·1. Let E be closed in the space R. Let f_0 and g_0 be two homotopic mappings of E into the n-sphere S^n. Then, if there exists an extension f of f_0 over R, there also exists an extension g of g_0 over R with f and g homotopic.

This is 'Borsuk's theorem' or 'Theorem VI. 5' of the book of Hurewicz & Wallman; their proof, given for the metric case, is literally applicable to our general case.

Theorems belonging to the topology of polytopes so far as they are needed in this paper are supposed to be known. Among these the attention of the reader is called to the invariance theorem concerning the Betti groups. The interpretation of this theorem for ∇-groups is as follows. Let C be a simplicial decomposition of the polytope $\bar{C}$. Then C is the nerve of the open covering σ of $\bar{C}$ whose elements are the open stars of C. Therefore, any ∇-cycle on C is an element of a ∇-class of $\bar{C}$. Conversely (and this is the meaning of the invariance theorem) each ∇-class of $\bar{C}$ contains ∇-cycles of the complex C, and two ∇-cycles of C are homologous on C to each other if and only if they belong to the same ∇-class of $\bar{C}$. Thus there is a fixed $(1, 1)$-correspondence between the elements of $\nabla^r_{\bar{C}}$ and those of ∇^r_C, this correspondence being an isomorphism.

Now let C be the nerve of a correct open covering ω of the bicompactum R. Let C be realized geometrically as a simplicial decomposition of a polytope $\bar{C}$ and consider any (ω, C)-mapping f of R into $\bar{C}$. Let us consider what is the homomorphism† H^*_f of $\nabla^r_{\bar{C}} = \nabla^r_C$ in ∇^r_R. To this end let $\mathfrak{z}^r$ be an element of $\nabla^r_{\bar{C}}$, and $z^r \in \mathfrak{z}^r$ a ∇-cycle on C.

---

† The homomorphisms H^*_f, J^R_E, etc., were defined in § 2.

Since the complex C is at the same time the nerve of σ and the nerve of ω, and f is an (ω, C)-mapping, the elements of ω are the inverse images under f of the elements of σ; therefore by the definition of the homomorphism H^*, $H^* \mathfrak{z}^r$ is nothing but the ∇-class of R to which z^r belongs as element. Thus the following has been proved:

LEMMA 5·2. Let f be an (ω, C)-mapping of the bicompactum R into the polytope $\bar{C}$, whose simplicial decomposition C is the nerve of ω. To obtain the homomorphism H_f^* of $\nabla^r_{\bar{C}}$ in ∇^r_R one has to take in any ∇-class $\mathfrak{z}^r \in \nabla^r_{\bar{C}}$ a ∇-cycle z^r of C; the ∇-class of R containing z^r is the element $H_f^* \mathfrak{z}^r$.

Proof of the Hopf extension theorem. (1) *Necessity.* Let f be an extension (over the bicompactum R) of the mapping f_0 of the closed set $E \subset R$ into the bicompactum R^*. Then on E

$$f_0(x) = f(J(x)),$$

where J is the identical mapping of E into R. This implies that for any element $z^* \in \nabla^r_E$

$$H^*_{f_0} z^* = J^R_E H^*_f z^*.$$

In other words: each element of ∇^r_E which is an image under $H^*_{f_0}$ (of an element z^* of $\nabla^r_{R^*}$) is an image under J^R_E (of the element $H^*_f z^* \in \Delta^r_R$), i.e. is extendable. Hence it follows, for $R = S^n$, that the degree of f_0 is extendable (to the degree of f).

(2) *Sufficiency.* Let $\dim R \leqslant n+1$, $E \subset R$ and let f be a mapping of E into S^n with degree $d \in \nabla^n_E$; suppose d extendable to $\mathfrak{z} \in \nabla^n_R$,

$$d = J^R_E \mathfrak{z}.$$

One can easily see that there exists an open covering

$$\alpha^* = \{O^*_1, \ldots, O^*_s\} \quad \text{of } R$$

with the following properties:

(*a*) There is a $\mathfrak{z}_{\alpha^*} \in \nabla^n_{\alpha^*}$ in $\mathfrak{z}$, i.e. $\mathfrak{z}_{\alpha^*} \subseteq \mathfrak{z}$.

(*b*) There is for any O^*_i at least one vertex e of (the fixed simplicial decomposition of) S^n such that $f(E \cap O^*_i)$ is contained in the open star Oe of e. Denote by e_i a fixed one among these vertices e.

(*c*) The covering α^* is correct and of dimension $\leqslant n+1$. Denote by α the covering of E whose elements are the (non-empty) intersections $E \cap O^*_i$ (remember that these are the inverse images of the O^*_i under the identical mapping I of E into R).

Let g_α be the simplicial mapping of C_α into the complex S^n determined by correlating the vertex e_i to $(E \cap O^*_i)$. Let $z_\alpha \in \nabla^n_\alpha$ be the degree of g_α and z_0 the ∇-class of E containing $\mathfrak{z}_\alpha$ (as subset). I say that $\mathfrak{z}_0 = d$. To prove this take the fundamental ∇-cycle x^n of S^n; define z^n_α on C_α by its values on the elements $t^n_\alpha \in C_\alpha$: if the vertices of t^n_α correspond to $(E \cap O^*_{i_0}), \ldots, (E \cap O^*_{i_r})$, we put

$$(z^n_\alpha . t^n_\alpha) = x^n(e_{i_0} \ldots e_{i_r}).$$

Obviously $z^n_\alpha \in z_\alpha$.

Now any $E \cap O^*_i$ is contained in $g_\alpha^{-1}(Oe_i)$; thus α is a refinement of $\omega = \{g_\alpha^{-1}(Oe_i)\}$ and $(E \cap O^*_i) \rightarrow g_\alpha^{-1}(Oe_i)$ is a projection. Therefore $z^n_\alpha \in d$ and hence $\mathfrak{z}_\alpha \subseteq d$.

Now
$$J^R_E\mathfrak{z} = d, \quad \mathfrak{z} \in \nabla^n_R, \quad d \in \nabla^n_E,$$
$$\mathfrak{z}_{\alpha^*} \subseteq \mathfrak{z}, \quad \mathfrak{z}_\alpha \subseteq d,$$
and $J^R_E\mathfrak{z}$ is by definition the ∇-class of E containing $J^{\alpha^*}_\alpha \mathfrak{z}_{\alpha^*}$. As $\mathfrak{z}_\alpha \in d = J^R_E\mathfrak{z}$, then (for an appropriate $\beta^* \geqslant \alpha^*$ and $\beta = E\beta^*$)
$$\pi^\alpha_\beta J^{\alpha^*}_\alpha \mathfrak{z}_{\alpha^*} = \pi^\alpha_\beta \mathfrak{z}_\alpha$$
or (as $J^{\alpha^*}_\alpha$ commutes with π^α_β)
$$J^{\beta^*}_\beta \pi^{\alpha^*}_{\beta^*} \mathfrak{z}_{\alpha^*} = \pi^\alpha_\beta \mathfrak{z}_\alpha.$$
Replacing our α by this β and setting again α for β it may be supposed that from the beginning
$$J^{\alpha^*}_\alpha \mathfrak{z}_{\alpha^*} = \mathfrak{z}_\alpha,$$
i.e. that $\mathfrak{z}_{\alpha^*}$ is an extension of $\mathfrak{z}_\alpha$.

Now take the polytopes $\bar{C}_{\alpha^*}$ and $\bar{C}_\alpha \subseteq C_{\alpha^*}$ and consider g_α as a mapping of $\bar{C}_\alpha$ into S^n. As $\mathfrak{z}_{\alpha^*}$ is an extension of $\mathfrak{z}_\alpha$, one has by the Hopf extension theorem for polytopes an extension g_{α^*} (over $\bar{C}_{\alpha^*}$) with degree $\mathfrak{z}_{\alpha^*}$ of the continuous mapping g_α.

Let b_{α^*} be an (α^*, C_{α^*})-mapping of R into $\bar{C}_{\alpha^*}$; then b_{α^*} by its very construction maps E into $\bar{C}_\alpha$ and
$$g(x) = g_\alpha(b_\alpha(x)) \quad (x \in E),$$
is a mapping of E into S^n;
$$g^*(x) = g_{\alpha^*}(b_{\alpha^*}(x)) \quad (x \in R)$$
is an extension of g over R.

The degree of g^* is the image of the degree $\mathfrak{z}_{\alpha^*} \in \nabla^n_{C_{\alpha^*}}$ of g_{α^*} under the homomorphism (of $\nabla^n_{C_{\alpha^*}}$ in ∇^n_R) induced by b_{α^*}. But in virtue of lemma 5·2 and of $\mathfrak{z}_{\alpha^*} \subseteq \mathfrak{z}$ this image is nothing but $\mathfrak{z}$. Moreover, one concludes easily from the definitions of g, b_{α^*} and g_α that for any $x \in E$ both $f(x)$ and $g(x)$ lie on some closed simplex of the complex S^n, whence it follows that f and g are homotopic. As g admits an extension over R, namely, g^*, the mapping f admits (by lemma 5·1) an extension f^* which is homotopic to g^* and thus has the degree z, Q.E.D.

From this theorem there follows as an easy corollary one of the assertions of the Hopf classification theorem, namely:

Corollary. Let the bicompactum R be of dimension $\leqslant n$. Corresponding to each element $\mathfrak{z}^n$ of ∇^n_R there exists a mapping of R in S^n with the degree $\mathfrak{z}^n$.

Proof. Denote by E a closed subset of R, containing one point only. Then the only element of ∇^n_E is the zero element; this element is of course extendable, and any element of ∇^n_R, in particular, the element $\mathfrak{z}^n$, is an extension of zero. Therefore, by the Hopf extension theorem, a mapping of E into S^n (which correlates to the unique point of E an arbitrary (but fixed) point of S^n) can be extended to a mapping f of R into S^n with the degree $\mathfrak{z}^n$.

The second part of the Hopf classification theorem is contained in theorem 2·20. There remains the third part:

Two mappings of the at most n-dimensional bicompactum R into S^n are homotopic if they have the same degree.

The proof of this statement follows from the Hopf extension theorem and theorem 4·7 (see Hurewicz & Wallman 1935, pp. 149–150).

We now proceed to the proof of

THEOREM 5·1. If the dimension of the bicompactum R is n, then R contains a canonical closed set E such that there exists a $\mathfrak{z}^{n-1} \in \nabla_E^{n-1}$ which is not extendable over R; on the other hand, for $r \geqslant n$ and an arbitrary choice of the closed set $E \subset R$ and of $\mathfrak{z}^r \in \Delta_E^r$, the element z^r is always extendable over R.

In other words:

Among all finite-dimensional bicompacta the n-dimensional ones are completely characterized by the property: the number $n-1$ is the greatest among those numbers r for which there exists a closed $E \subset R$ and a $\mathfrak{z}^r \in \nabla_E^r$ non-extendable over R.

The proof is evident: By theorem 4·50 there is a mapping f of a closed $E \subset R$ into S^{n-1} which is not extendable over R; by the Brouwer-Urysohn extension theorem, E can be supposed canonical; the degree of the mapping f is a non-extendable element of ∇_E^{n-1}. On the other hand, let a closed $E \subset R$ and $\mathfrak{z}^r \in \nabla_E^r$ $(r \geqslant n)$, be chosen arbitrarily. Take the confinal part of the spectrum of the open coverings of R consisting of all n-dimensional coverings and consider among the $\mathfrak{z}_\alpha^r \in z^r$ only those which belong to the nerves C of the elements of this confinal part. Since, for $r > n$, all these z_α^r are zero, it follows that $\mathfrak{z}^r = 0$, and $\mathfrak{z}^r$ is extendable. For $r = n$, the z_α^n (being n-dimensional chains on the n-dimensional complexes C_α), are ∇-cycles of the complexes C_α, and $\mathfrak{z}^n$ is again extendable.

We shall now put theorem 5·1 in some different forms. For this purpose notations introduced in Alexandroff (1943) will be used. Hence

If R is an n-dimensional bicompactum, then there exists a closed $E \subset R$ such that $\nabla_{ER}^{n-1} \neq \nabla_E^{n-1}$, i.e. $\nabla_{E:R}^{n-1} \neq 0$, or, what is the same, $\Delta_{E:R}^{n-1} \neq 0$, while, on the other hand, for any closed $E \subset R$ and $r \geqslant n$ $\nabla_{E:R}^r = 0$ $(\Delta_{E:R}^r = 0)$.

The 'Δ-form' of the last theorem means:

There is, in the n-dimensional bicompactum R, a closed $E \subset R$ and a thread $\zeta^{n-1} = \{\zeta_\alpha^{n-1}\} \in \Delta_E^{n-1}$ which 'does not bound on E' (in the sense that ζ^{n-1} is not the zero element of Δ_E^{n-1}), but 'bounds on R' in the sense that if the covering α of E has the form $\alpha = E\alpha^*$, α^* being a covering of R, each of the cycles $z_\alpha^{n-1} \in \zeta_\alpha^{n-1}$, bounds on α^*. On the other hand, there is, for $r \geqslant n$, no $\zeta^r \in \Delta_E^r$ which is different from zero and bounds on R (and, as follows from the proof of theorem 5·1, for $r > n$, no ζ^r different from zero at all).

In these theorems the Betti groups may be defined by means of any of the spectra (1)–(5) of § 2 or even by means of a confinal part of any of them. In particular, we can use for R the exact spectrum Σ and E for the transitive spectrum $E\Sigma$ whose elements are the nerves of the coverings α of E of the form $\alpha = E\alpha^*$, where α^* is an exact covering of R. This allows to speak about the true cycles on R and on E respectively: a true cycle on R (on E) is a thread $z^r = \{z_\alpha^r\}$, where the z_α^r are cycles on α^* (on $E\alpha^*$). The cycle $z^r = \{z_\alpha^r\}$ on E bounds (on E) if each z_α^r bounds on $E\alpha^*$; the cycle $z^r = \{z_\alpha^r\}$ (on E) bounds on R, if each z_α^r bounds on α^*. These definitions allow us to give the following form to theorem 5·1 which exactly corresponds to the original form of the fundamental dimension theorem as given in my paper (1932) for compacta ('2. Hauptsatz' on p. 195 of that paper):

An n-dimensional bicompactum R necessarily contains a closed set E and on it an $(n-1)$-dimensional true cycle which bounds on R but does not bound on E. On the other hand, for any $r \geqslant n$ and any closed $E \subset R$, each r-dimensional true cycle on E bounds on E, provided that it bounds on R.

The third pair of isomorphisms of my paper (1943) will now be applied, i.e. the isomorphisms

$$\nabla^r_{E:R} \approx \nabla^{r+1}_{(R \setminus E):R}, \quad \Delta^r_{E:R} \approx \Delta^{r+1}_{(R \setminus E):R}.$$

Let R be an n-dimensional bicompactum. According to theorem 5·1 choose the closed $E \subset R$ so as to have $\nabla^{n-1}_{E:R} \neq 0$. Then by the above isomorphism $\nabla^n_{(R \setminus E):R} \neq 0$, and since $\nabla^n_{(R \setminus E):R} \subseteq \nabla^n_{R \setminus E}$ *a fortiori* $\nabla^n_{R \setminus E} \neq 0$. Dually $\Delta^n_{R \setminus E} \neq 0$.

Taking a confinal part made up of n-dimensional complexes (e.g. in the spectrum of the open coverings of R), one sees at once that for $r > n$ and any choice of the closed set $E \subset R$ the groups $\nabla^r_{R \setminus E}$, $\Delta^r_{R \setminus E}$ are zero. Thus theorem 4·2 is completely proved in both its (∇- and Δ-) forms.

Remark. In all formulations of theorem 5·1 as well as in theorem 4·2 the closed set E can be supposed canonical.

6. NORMAL SPACES. PROOF OF THEOREM 4·1

A bicompactum bR is called an extension of the space R if R is an everywhere dense subset of bR, i.e.

$$R \subseteq bR, \quad (bR)(R) = bR.$$

Among the bicompact extensions of R there is one called the Čech extension or the *maximal extension* βR. This extension is completely characterized by the property: βR can be mapped on to any bR in such a way that every point of R is a fixed point of this mapping (and that no point of $\beta R \setminus R$ is carried into a point of R).

I have shown (Alexandroff 1939; see also Fomin 1943) that βR may be obtained from R by the following very intuitive construction:

Construction of βR. A system S of subsets M_α of a given set M is called centred if any finite number of the sets M_α has a non-empty intersection. A system S of open sets H_α of the space R is called regular if to each $H_\alpha \in S$ an $H_\beta \in S$ may be found in such a way that $RH_\beta \subseteq H_\alpha$. A centred regular system S of open sets $H_\alpha \subseteq R$ is called an end (of the space R) if there is no centred regular system (of open sets of R) different from S and containing S as a subsystem. The ends $x = \{H_\alpha\}$ of R are by definition the points of βR; to get the inclusion $R \subseteq \beta R$ any point $x \in R$ will be identified with the end composed by all open sets of R containing the point x. The open sets H_α which are elements of an end x are called the coordinates of the point $x \in R$.

To each open $H \subseteq R$ there corresponds the set $O_H \subseteq \beta R$ of all points $x \in \beta R$ such that $H \in x$ (in other words: O_H is the set of all points $x \in \beta R$ having H among their coordinates). Open sets in βR are defined as the sets O_H, together with all their sums. In this way βR is topologized. The same topology may obviously be obtained by defining βR as a 'neighbourhood space', where a neighbourhood of the point $x = \{H_\alpha\}$ is any set of the form O_{H_α}.

The space βR can also be defined by the following construction due to Wallman (1938). Points of βR are defined as maximal centred systems of closed sets $\phi \subseteq R$,

$$x = \{\phi_\alpha\}$$

('maximal' means again that x is not a proper subsystem of a centred system of closed sets of R). A point $x \in R$ is in this definition identified with the system of all closed $\phi \subseteq R$ containing the point x. In this way again $R \subseteq \beta R$. If ϕ is any closed set in R then E_ϕ is by definition the set of all $x \in \beta R$ with $\phi \in x$. Obviously $\phi \subseteq E_\phi$. The topology in βR is now introduced by the definition of closed sets in βR as the sets of the form E_ϕ (with ϕ closed in R), together with all intersections of such sets.

THEOREM 6·10. If we put in correspondence with each canonical closed set A of the space R its closure $(\beta R) A$ in βR, a (1:1)-correspondence between the system of all canonical closed sets of R and the system of all canonical closed sets of βR is obtained; this (1:1)-correspondence is an intersection-isomorphism, i.e.

(1) $A_1 \cap \ldots \cap A_r = \Lambda$ implies $(\beta R) A_1 \cap \ldots \cap (\beta R) A_r = \Lambda$, and

(2) $(\beta R) A_1 \subseteq (\beta R) A_2$ implies $A_1 \subseteq A_2$.

Proof. We first prove (1). Let $E_i = E_{A_i}$. Since $E_i \supseteq A_i$ is closed, then $A_i \subseteq (\beta R) A_i \subseteq E_i$; thus (1) is contained in

(3) '$A_1 \cap \ldots \cap A_r = \Lambda$ implies $E_1 \cap \ldots \cap E_r = \Lambda$.'

Now $x \in E_1 \cap \ldots \cap E_r$ means that $A_1, \ldots, A_r$ are among the co-ordinates of x, thus $A_1 \cap \ldots \cap A_r \neq \Lambda$.

To prove (2) suppose that $A_1 \setminus A_2$ contains a point a_1. The closed sets a_1 and A_2 (in R) are disjoint; thus the same is true for $a_1 = (\beta R) a_1$ and $(\beta R) A_2$, i.e. the point $a_1 \in A_1 \subseteq (\beta R) A_1$ does not belong to $(\beta R) A_2$, and (2) is established. By the same argument, to any distinct A_1, A_2 there correspond distinct $(\beta R) A_1$, $(\beta R) A_2$.

The hypothesis that the sets A are canonical has not yet been used: we have simply proved that the operator (βR) maps isomorphically the system of all closed sets of R into the system of all closed sets of βR. It remains to show that:

(1) The operator (βR) transforms canonical closed sets of R into canonical closed sets of βR.

(2) Any canonical closed set of βR is the image under (βR) of a canonical closed set of R.

Proof of (1). LEMMA. Every open $H \subseteq R$ is dense in the corresponding O_H.

In fact, let $x = \{\ldots H \ldots\} \in O_H$. Take any neighbourhood $O_{H'}$ of x. From $H' \in x$ it follows that $H \cap H' \neq \Lambda$; the points of $H \cap H'$ being points of H belonging to $O_{H'}$, the lemma is proved.

Now let A be canonical closed in R; this means that $A = R(U)$. Since U is dense in O_U,

$$(\beta R)\,U = (\beta R)\,O_U.$$

But $(\beta R)\,U = (\beta R)\,(RU) = (\beta R)\,A$; thus $(\beta R)\,A = (\beta R)\,O_U$ is canonical.

Proof of (2). Let $A = (\beta R)\,U$ be canonical closed in βR; suppose that U is open canonical in βR. Let

$$u = U \cap R.$$

Since u is dense in U, then

$$A = (\beta R)\,U = (\beta R)\,u = (\beta R)\,(Ru);$$

as Ru is canonical in R, the assertion (2) and thus theorem 6·10 are completely proved.

Theorem 6·10 may be formulated as follows:

THEOREM 6·1. The normal space R and its maximal extension have the same spectrum.

Since the spectrum of a bicompactum is a Hausdorff spectrum (Kurosch 1935), there follows

COROLLARY 6·1. The spectrum of a normal space is a Hausdorff spectrum.

The space defined by the spectrum of a bicompactum is (as shown by Kurosch 1935) this bicompactum; this, together with theorem 6·1, gives a new and important construction for the maximal extension of a space:

THEOREM 6·2. The maximal extension of a space R is the bicompactum defined by the spectrum of the space R.

THEOREM 6·3. The operator O_U applied to the canonical open sets U of the space R realizes a $(1:1)$-correspondence between the system of all canonical open sets of R and the system of all canonical open sets of βR; this correspondence is an intersection isomorphism:

(4) $O_{U_1} \cap \ldots \cap O_{U_r} = \Lambda$ if and only if $U_1 \cap \ldots \cap U_2 = \Lambda$.

(5) $O_{U_1} \subseteq O_{U_2}$ if and only if $U_1 \subseteq U_2$.

Proof. LEMMA 2. For any canonical open set $U \subseteq R$:

(6) $O_U = \beta R \setminus (\beta R)\,(R \setminus RU)$.

Proof of Lemma 2. Let $x \in O_U$. Then $U \in x$. Now $(\beta R)\,(R \setminus RU)$ is the set of all $x' = \{H_\alpha\}$ whose coordinates all intersect $R \setminus RU$; therefore $x \in \beta R \setminus (\beta R)\,(R \setminus RU)$. Thus, the inclusion $O_U \subseteq \beta R \setminus (\beta R)\,(R \setminus RU)$ is proved. Let $x \in \beta R \setminus (\beta R)\,(R \setminus RU)$; this implies the existence of an $H \in x$ not intersecting $R \setminus RU$, i.e. $H \subseteq JRU = U$, since U is canonical. From $H \subseteq U$, $H \in x$ it follows that $U \in x$, $x \in O_U$, which proves the inclusion $\beta R \setminus (\beta R)\,(R \setminus RU) \subseteq O_U$ and with it lemma 2.

Remark. For canonical U the set $R \setminus RU$ is dense in $R \setminus U$, therefore

$$(\beta R)\,(R \setminus RU) = (\beta R)\,(R \setminus U) \quad \text{and} \quad O_U = \beta R \setminus (\beta R)\,(R \setminus U).$$

Since the canonical open and the canonical closed sets are complementary to each other, it follows from lemma 2 together with theorem 6·10 that the operator O_U realizes a (1 : 1)-correspondence between the canonical open sets of R and those of βR. In order to prove the isomorphic character of this correspondence we first recall (see Alexandroff 1939) that

$$O_{U_1}\cap\ldots\cap O_{U_r}=O_{U_1\cap\ldots\cap U_r}.$$

Furthermore, from $\quad O_{U_1}\subseteq O_{U_2},$

there follow $\quad \beta R\setminus(\beta R)(R\setminus RU_1)\subseteq\beta R\setminus(\beta R)(R\setminus RU_2)$

and $\quad (\beta R)(R\setminus RU_1)\supseteq(\beta R)(R\setminus RU_2),$

or (by the last Remark) $\quad (\beta R)(R\setminus U_1)\supseteq(\beta R)(R\setminus U_2).$

Thus, since $R\setminus U_1$ and $R\setminus U_2$ are canonical closed,

$$R\setminus U_1\supseteq R\setminus U_2,\quad U_1\subseteq U_2.$$

This completes the proof of theorem 6·3.

As an immediate consequence of theorem 6·3 there follows

COROLLARY 6·3: $\quad \dim\beta R=\dim R.$

(This result is known—see Wallman (1938), footnote 11 on p. 119.)

Since for spaces with the first countability axiom homeomorphism between βR_1 and βR_2 implies homeomorphism between R_1 and R_2 we deduce from theorem 6·10:

THEOREM 6·11. Any normal space R satisfying the first countability axiom is completely determined by the combinatorial structure of the system of its canonical closed sets.

More precisely: Suppose that there is a multiplicative isomorphism between the system of all canonical closed sets of R_1 and the system of all canonical closed sets of R_2; then the spaces R_1 and R_2 are homeomorphic.

I do not know whether the analogous theorem is true for canonical open sets. Moreover, it is still unknown whether theorem 6·11 remains valid without the assumption of the first countability axiom.

The following Remark is evident. Let $S=\{H_\alpha\}$ and $S'=\{H'_\alpha\}$ be two systems of open sets, of the spaces R and R' respectively; a (1 : 1)-correspondence between S and S' is called for the moment a regularity-isomorphism, if it preserves relations of the form

$$H_1\cap\ldots\cap H_r=\Lambda,\quad H'_1\cap\ldots\cap H'_r=\Lambda,$$

$$R(H_1)\subseteq H_2,\quad R'(H'_1)\subseteq H'_2.$$

Now suppose that there exists a regularity-isomorphism between the system of all open sets of R and the system of all open sets of R'. Since a regularity-isomorphism transforms in a (1 : 1)-manner the ends of the one space into the ends of the other a (1 : 1)-mapping of βR on to $\beta R'$ is induced. Furthermore, this mapping transforms sets of the form $O_H\subseteq\beta R$ into sets of the form $O_{H'}\subseteq\beta R'$ and vice versa, and is therefore a homeomorphism between βR and $\beta R'$. If R and R' satisfy the first countability axiom, a homeomorphism between βR and $\beta R'$ implies a homeomorphism between R and R'. Moreover, it is easily seen that any end is completely determined by its canonical elements, and that open sets of the form O_U (U canonical) are sufficient for the definition of the topology in βR. Thus one obtains:

THEOREM 6·4. Let the spaces R, R' satisfy the first countability axiom. Denote by $\mathbf{O}$, $\mathbf{U}$, $\mathbf{O}'$, $\mathbf{U}'$ the system of all open sets and of all open canonical sets of each of these spaces. If there exists a regularity-isomorphism between $\mathbf{O}$ and $\mathbf{O}'$ or between $\mathbf{U}$ and $\mathbf{U}'$, the spaces R and R' are homeomorphic.

Proof of Theorem 4·1. Let $\dim R = n$. Then $\dim \beta R = n$; thus by theorem 4·2 for bicompacta the spectrum of βR, i.e. the spectrum of R, has the structure described in theorem 4·1, Q.E.D.

Proof of Theorem 4·2 (for a normal R). Let $\dim R = \dim \beta R = n$. Then there exists in βR a canonical closed set A such that $\nabla^n(\beta R \bmod A)$ is different from zero. If we define the last group by means of the spectrum of the open canonical coverings of βR and write (in the notations of §§ 1 and 2)

$$\nabla^n(\beta R \bmod A) = \lim\,(\nabla^n_{\alpha 1}, \pi^{\alpha 1}_{\beta 1})$$

with $\nabla^n_{\alpha 1} = \nabla^n_{H_\alpha}$, $H_\alpha = C_\alpha \setminus A_\alpha$, $A_\alpha = AC_\alpha$, it is seen that the C_α may be regarded as nerves of the open canonical coverings of R as well as of βR. Let furthermore $A = (\beta R)U$, $R \cap U = U'$, $A' = RU'$. Then $A_\alpha = A'C_\alpha$ (in the spectrum C of the open canonical coverings of R). Therefore the above group $\lim\,(\nabla^n_{\alpha 1}, \pi^{\alpha 1}_{\beta 1})$ is at the same time the group $\nabla^n(R \bmod A')$. Since $\lim\,(\nabla^n_{\alpha 1}, \pi^{\alpha 1}_{\beta 1}) = \nabla^n(R \bmod A')$ is seen to be different from zero, theorem 4·2 is proved in its full generality.

7. CANTOR MANIFOLDS

Let $C = \{C_\alpha, \varpi^\beta_\alpha\}$ be a Hausdorff spectrum. This spectrum is called n-dimensional if it satisfies the following condition:

There exists a maximal subspectrum A of C such that the n-dimensional Betti group of $C \bmod A$ is different from zero, while for $r > n$ and for any choice of the maximal subspectrum A of C the r-dimensional Betti group of $C \bmod A$ is the zero group.

In this definition the infinite cyclic group J is taken as the coefficient-group for the ∇-case. For the Δ-case one has, of course, to take as coefficient-group the group J of all real numbers reduced mod 1; one can define in the same manner the dimension of a spectrum with respect to any coefficient-group.

A maximal subspectrum A of the n-dimensional spectrum C is called *an* (*abstract*) *Cantor manifold lying on* C, if, whatever be the representation of A as the sum of two maximal subspectra (no one of which coincides with A), the intersection of these subspectra is at least $(n-1)$-dimensional. If, in particular, C is the spectrum of the normal space R then a Cantor manifold lying on C is called an *abstract Cantor manifold lying on the n-dimensional space R*.

Remark 1. It is unknown if a Cantor manifold lying on an n-dimensional spectrum (in particular on an n-dimensional space) is itself an n-dimensional spectrum. This seems to me to be an outstanding problem in the general dimension theory of spectra.

In conformity with the well-known definition given by Urysohn more than twenty years ago, let us now define a *classical Cantor manifold lying on an n-dimensional bicompactum R* as a closed set $E \subseteq R$ having the property that for any representation of E as the sum of two closed proper subsets, the intersection of these subsets is at least $(n-1)$-dimensional.

Remark 2. It is again unknown whether a classical Cantor manifold lying on an n-dimensional bicompactum R is itself an n-dimensional bicompactum: in the metric case the dimension of R (defined as in § 4 by means of coverings) is equal to the *inductive dimension* $\operatorname{ind} R$ (defined by the well-known Urysohn-Menger process of induction beginning with $\dim \Lambda = -1$ for the empty set Λ). Now in the case of a bicompactum R it is only known that

$$\operatorname{ind} R \geqslant \dim R$$

(as I proved in 1941*b*), while it still remains unknown whether the inequality $\operatorname{ind} R \leqslant \dim R$ holds. Here again is an important problem of general dimension theory.

It follows immediately from what was said at the end of § 3 that for an n-dimensional bicompactum there is a $(1:1)$-correspondence between the abstract Cantor manifolds lying on R and the classical ones. The purpose of this section is to prove the following:

THEOREM 7·1. On any n-dimensional normal space (n being finite), there lie abstract Cantor manifolds.

This theorem is first proved for a bicompact R. The proof will closely follow the proof given by Hurewicz & Wallman for the metric case.

LEMMA 7·1. Let f and g be two mappings of the bicompactum R into the sphere S^{n-1}. Suppose that the points in which $g(x) \neq f(x)$ form a subset $M \subseteq R$ of a relative dimension $\leqslant n-2$. Then f and g are homotopic.

Proof. Following Hurewicz & Wallman, we define the closed set E in $[R, I]$ consisting of all points $(x, 0)$ and $(x, 1)$ with $x \in R$, and all points (x, t) with $x \in R \setminus M$ and $0 \leqslant t \leqslant 1$. (The set E is closed because M is obviously open.) The mapping $f(x, t)$ of E into S^{n-1} is defined as follows:

$$\begin{cases} f(x,t) = f(x) = g(x) \quad \text{for} \quad x \in R \setminus M, \\ f(x,0) = f(x), \quad f(x,1) = g(x). \end{cases}$$

The set $[R, I] \setminus E$ is contained in $[M, I]$. Thus by the corollary to theorem 4·7:

$$\operatorname{rd}([R, I] \setminus E) \leqslant \operatorname{rd}[M, I] < (n-2)+1 = n-1.$$

Therefore, by theorem 4·6, the mapping $f(x, t)$ defined on E can be extended over $[R, I]$, and this proves the homotopy of g and f.

LEMMA 7·2. Suppose that the bicompactum R is the sum of two closed subsets E_1 and E_2. Suppose that f_1, f_2 are mappings into S^{n-1} of E_1, E_2 respectively, and that the points (in $E_1 \cap E_2$) in which $f_1(x) \neq f_2(x)$ form a set of relative dimension $\leqslant n-2$. Then f_1 can be extended over R.

Proof. The mappings f_1 and f_2 considered as mappings of the bicompactum $E_1 \cap E_2$ are homotopic by lemma 7·1. Since f_2 (as a mapping of $E_1 \cap E_2$ into S^{n-1}) is extendable over E_2 (to the mapping f_2) it follows from the Borsuk lemma 5·1 that there is an extension f_{12} of f_1 (considered on $E_1 \cap E_2$) over E_2. Defining

$$\begin{cases} f(x) = f_1(x) & \text{on} \quad E_1 \\ f(x) = f_{12}(x) & \text{on} \quad E_2 \end{cases}$$

we obtain the desired extension of f_1 over R is obtained.

LEMMA 7·3. Let there be given a bicompactum R, a closed set $E \subset R$ and a mapping f of E into a sphere S. Suppose, furthermore, that there is in R a decreasing well-ordered system of closed sets E_λ (the λ are finite or transfinite ordinals). Suppose that f admits no extension over a set of the form $E \cup E_\lambda$; then f admits no extension over $E \cup E'$, where $E' = \bigcap_\lambda E_\lambda$.

The easy proof of this lemma is as follows. Suppose that f is extendable over $E \cup E'$. Then by the Brouwer-Urysohn extension theorem f can be extended over a certain neighbourhood U of $E \cup E'$ in R. Since R is bicompact, all E_λ with a sufficiently great λ are contained in U, thus f would be extendable over some $E \cup E_\lambda$, contradicting our hypothesis.

From lemma 7·3 follows:

LEMMA 7·4. Given a bicompactum R, a closed set $E \subset R$ and a mapping f of E in a sphere S which cannot be extended over R, there is a closed $E' \subseteq R$ such that f cannot be extended over $E \cup E'$ but, for any $E'' \subset E'$, admits an extension over $E \cup E''$.

It is now easy to prove theorem 7·1 for a bicompactum R.

Let $\dim R = n$. Then there exists a closed $E \subset R$ and a mapping f of E into S^{n-1} which is non-extendable over R. By lemma 7·4 there is in R a closed E' such that f is extendable over $E \cup E''$ if $E'' \subset E'$, but is not extendable over $E \cup E'$. I say with Hurewicz & Wallman that this E' is a classical Cantor manifold on the bicompactum R. Otherwise, let

$$E' = E_1' \cup E_2', \quad E_1' \neq E' \neq E_2', \quad \dim (E_1' \cap E_2') \leqslant n-2.$$

By the definition of E' the mapping f can be extended over $E_i = E \cup E_i'$ $(i = 1, 2)$, to a mapping f_i, and the points in which f_1 possibly differs from f_2 are all in $E_1' \cap E_2'$ and form therefore a point set in $E_1 \cap E_2$ of relative dimension not exceeding $n-2$ (with respect to the bicompactum $E \cup E' = E_1 \cup E_2$). It follows by lemma 7·2 that f_1, and thus f, admits an extension over $E \cup E'$, contradicting the choice of E'.

Now let R be an arbitrary n-dimensional normal space. Its spectrum, being at the same time the spectrum of an n-dimensional bicompactum, namely, of βR, contains a Cantor manifold which by its definition lies on R. Thus theorem 7·1 is proved in its full generality.

REFERENCES

Alexandroff, P. 1932 *Math. Ann.* **106**, 161–238.
Alexandroff, P. 1939 *Math. Sbornik.* **5**, 403–423.
Alexandroff, P. 1940 *C.R. Acad. Sci. U.R.S.S.* **26**, 619–622.
Alexandroff, P. 1941*a* *Trans. Amer. Math. Soc.* **49**, 41–105.
Alexandroff, P. 1941*b* *Mitt. Akad. Wiss. Georgischen S.S.R.* **2**, 1–6.
Alexandroff, P. 1943 *Trans. Amer. Math. Soc.* **54**, 286–339.
Alexandroff, P. & Hopf, H. 1935 *Topologie.* Berlin: Springer; also New York: Dover Publ.
Alexandroff, P., Hopf, H. & Pontrjagin, L. 1935 *Comp. math.* **4**, 239–255.
Fomin, S. 1943 *Ann. Math.* **44**, 471–480.
Hurewicz, W. & Wallman, H. 1941 *Dimension theory.* Princeton Univ. Press.
Kurosch, A. 1935 *Comp. math.* **2**, 471–476.
Lefschetz, S. 1942 *Algebraic topology.* New York: Amer. Math. Soc.
Wallman, H. 1938 *Ann. Math.* **39**, 112–126.

The crystal structures of two polyamides ('nylons')

By C. W. Bunn and E. V. Garner

Research Department, I.C.I. Ltd., Alkali Division, Northwich, Cheshire

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 6 *May* 1946
—*Read* 20 *March* 1947)

[Plates 1 and 2]

Detailed interpretations of the X-ray diffraction patterns of fibres and sheets of 66 and 6.10 polyamides (polyhexamethylene adipamide and sebacamide respectively) are proposed. The crystal structures of the two substances are completely analogous.

Fibres of these two polyamides usually contain two different crystalline forms, α and β, which are different packings of geometrically similar molecules; most fibres consist chiefly of the α form. In the case of the 66 polymer, fibres have been obtained in which there is no detectable proportion of the β form.

Unit cell dimensions and the indices of reflexions for the α form were determined by trial, using normal fibre photographs, and were checked by using doubly oriented sheets set at different angles to the X-ray beam. The unit cell of the α form is triclinic, with $a = 4{\cdot}9$ A, $b = 5{\cdot}4$ A, c (fibre axis) $= 17{\cdot}2$ A, $\alpha = 48\frac{1}{2}^\circ$, $\beta = 77^\circ$, $\gamma = 63\frac{1}{2}^\circ$ for the 66 polymer; $a = 4{\cdot}95$ A, $b = 5{\cdot}4$ A, c (fibre axis) $= 22{\cdot}4$ A, $\alpha = 49^\circ$, $\beta = 76\frac{1}{2}^\circ$, $\gamma = 63\frac{1}{2}^\circ$ for the 6.10 polymer. One chain molecule passes through the cell in both cases.

Atomic coordinates in α crystals were determined by interpretation of the relative intensities of the reflexions. The chains are planar or very nearly so; the oxygen atoms appear to lie a little off the plane of the chain. The molecules are linked by hydrogen bonds between C═O and NH groups, to form sheets. A simple packing of these sheets of molecules gives the α arrangement.

An alternative packing of the sheets gives a two-molecule triclinic cell, and this is the structure proposed for the β form.

Streaks on the layer lines (but not on the equator) of fibre photographs are attributed to an arrangement of molecular sheets in which α and β groupings occur at random.

The optical properties of fibres and doubly oriented sheets are consistent with the structure determined by the X-ray method. Unoriented specimens are spherulitic; optical evidence on the structure of the spherulites is discussed.

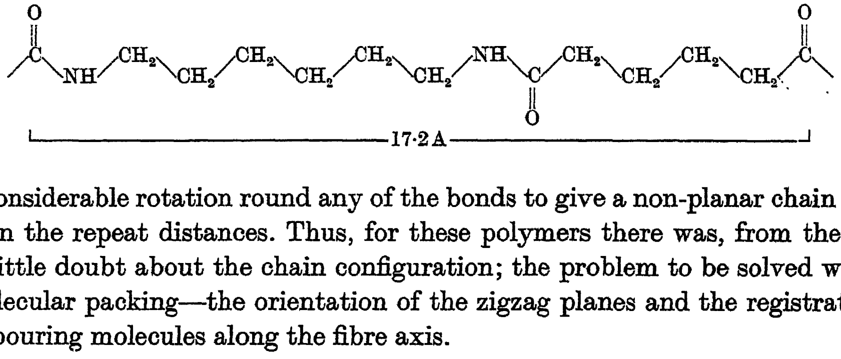
Some striking similarities between the crystallography of the polyamides and that of the fibrous proteins are discussed.

Polyamides of high molecular weight have attained considerable practical importance as 'nylon' fibres. The substance most widely used in this way is the polyamide made by the condensation of hexamethylene diamine, $NH_2(CH_2)_6NH_2$, and adipic acid, $HOOC(CH_2)_4COOH$, and known as '66' polymer. (In the abbreviated nomenclature for the polyamides made from diamines and dibasic acids, two numbers are given, the first being the number of carbon atoms in the diamine and the second the number of carbon atoms in the dibasic acid. Thus, the polymer made from hexamethylene diamine and sebacic acid, $HOOC(CH_2)_8COOH$, is known as '6.10'). In the attempt to correlate the physical properties with molecular structure, interest centres on the high tensile strength of the fibres (which in this respect resemble natural silk) and on the high melting-points of the polyamides as compared with polyethylene (66 nylon melts at 250° C, polyethylene at 115° C). It is natural to suppose that these features are due in some way to forces between the polar groups in the molecules—one expects to find hydrogen bonds between the C═O groups of one molecule and the NH groups of its neighbours; this expectation has been realized in the present work, in which O—N distances characteristic of hydrogen bonds have been found. The polyamides are also interesting on account of their relationship to the proteins: the grouping —CO—NH— is present in both types of substances, and any information on molecular packing gained from a detailed study of the structures of polyamides may be useful in dealing with the far more complex problems presented by the fibrous proteins. The expectation of finding some structural features common to polyamides and fibrous proteins has, in fact, been strikingly justified in the present work, which was finished in 1941.

The two polymers studied here by X-ray methods are those already mentioned—the 66 and 6.10 polyamides. It seemed likely from the start that the crystal structures of these two substances were analogous, and it soon became evident that the detailed interpretation of the X-ray diffraction pattern of either polymer by itself was likely to be very difficult. Both substances were, therefore, studied concurrently; hypothetical interpretations of particular features were applied to both, and correlation of the results of such tests not only eased the task of interpretation, but led to increased confidence in the results. It is doubtful whether, if only one polymer had been studied, it would have been possible to get as far as a detailed solution of the atomic structure such as those proposed in this paper; and certainly, any proposed interpretation would have been regarded with far more suspicion. The difficulties and ambiguities of the interpretation of X-ray diffraction patterns of high polymers are such that this procedure—examination of two or more structures which promise to be analogous—is of great value, and should be followed wherever possible.

Chain configuration. All the X-ray diffraction photographs used in the present work were taken by copper $K\alpha$ radiation. Those used for the detailed interpretation in terms of atomic structure were taken on cylindrical films, in order to obtain as wide an angular range of reflexions as possible. Drawn fibres of the 66 and 6.10 polymers give diffraction patterns showing well-marked layer lines; for structure determination, it is necessary that the crystals in the fibres should be as large and perfect as possible, giving sharp reflexions; and the orientation should be as perfect as possible, so that the arcs are as short as possible. Fuller, Baker & Pape (1940) have shown how this can be achieved by suitable thermal treatment. One of the best photographs obtained is shown in figure 6, plate 1; those of the 6.10 polymer were not quite so good as those of 66.

The repeat distances for the two polymers along the fibre axis, given approximately by the layer line spacings but determined more accurately later on in another way, are 17·2 A for the 66, and 22·4 A for the 6.10 polymer. It has already been pointed out by Fuller (1940) that, if it is assumed that the long molecules are parallel to the fibre axis, these figures suggest that the chains must be fully extended, or very nearly so: they must be plane or very nearly plane zigzags:



Any considerable rotation round any of the bonds to give a non-planar chain would shorten the repeat distances. Thus, for these polymers there was, from the start, very little doubt about the chain configuration; the problem to be solved was one of molecular packing—the orientation of the zigzag planes and the registration of neighbouring molecules along the fibre axis.

Unit cell. One dimension of the unit cell, which will be called the c axis (the 'height' of the cell), is the same as the molecular repeat distance just mentioned—17·2 A for the 66 polymer and 22·4 A for the 6.10. The dimensions of the base of the cell, projected on to a plane normal to the c axis, are obtained from the positions of the spots on the equators of the photographs. It is important to note that both the positions and the intensities of these equatorial reflexions are almost exactly the same for the two polymers, a fact which suggests that the molecules are packed side by side in almost exactly the same way in the two polymers. The simplest projected cell-base which fits the positions of the equatorial reflexions has edges 4·03 and 4·77 A[*] with an angle of 66° 15′ between them (*ABCD* in figure 1); there can be only one chain molecule passing through this cell, for the distance between unbranched chain molecules packed side by side is always of the order of 4 A.

---

[*] The revised X-ray wave-lengths given by Lipson & Riley (1943) have been used. Spacings given are in Ångström units, *not* kX units.

The first check of unit-cell dimensions is given by a comparison of calculated with observed density; and it can be made at this stage (that is, before the other angles of the unit cell are considered), since the volume of the cell is equal to the product of the height and the area of the projected cell-base. The volume of the cell is, for the 66 polymer, 301×10^{-24} c.c., the weight of one monomer unit 373×10^{-24} g., hence the density would be 1·24. The measured density of drawn fibres of 66 polymer is 1·14 to 1·16. The measured densities of crystalline polymers are usually lower by 5 to 10 % than the true densities of the crystalline regions, owing to the presence of amorphous material of lower density; hence the density of the crystalline regions of 66 polymer would be expected to be about 1·20 to 1·25; the density calculated for the proposed unit-cell dimensions is in satisfactory agreement with this figure.

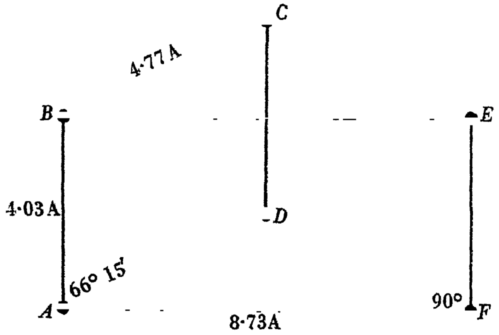
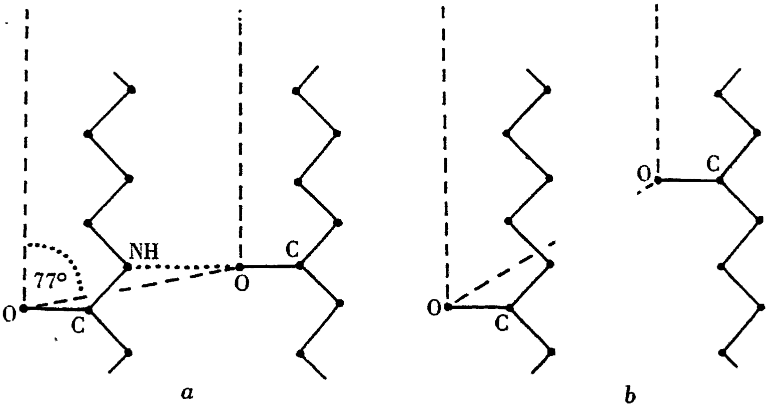


FIGURE 1. Alternative cell-bases projected on a plane perpendicular to *c* (fibre axis).

To find the other angles of the cell, it is necessary to consider the positions of all the other spots on the photographs; and in doing this, it is necessary to take into account the possibility that the true projected cell-base might be larger than that which fits the equatorial spots by themselves; in particular, it happens that the two-molecule cell *ABEF* in figure 1 is rectangular within the limits of error of measurement, and this may mean that the cell is an orthorhombic or monoclinic two-molecule cell which is apparently centred when seen along the *c* axis.

These possibilities, and many more, were tested, using reciprocal lattice rotation diagrams (see Bernal 1926), but it was found impossible to account for the whole of the spots by any simple orthorhombic or monoclinic cell for either of the two polymers; only improbably large cells having these symmetries would account for all the spots. Triclinic cells were then considered; and in doing this, attention was first confined to the 7th layer line of the 66 polymer. The reflexions on the 7th layer line are much stronger than those on other layers (apart from the equator), and there is thus a subperiod of $c/7$; this is evidently due to the fact that there are in the repeating unit seven zigzags, which to a first approximation can be regarded as equivalent. The procedure of confining attention to the 7th layer line is equivalent to reducing the repeat distance of the molecules to $c/7$ and considering the side-by-side packing of simple zigzag chains. Moreover, the positions and relative intensities of the reflexions on the 9th layer line of the photographs of 6.10 polymer are very similar to those of the 7th layer line of 66 polymer—a fact which is further evidence that the packing of the zigzags is the same in the two polymers, and which means that the procedure just mentioned should lead to a small pseudo-unit cell valid for both substances. It was found that the positions of these reflexions could be explained by a one-molecule triclinic cell.

The reflexions on the remaining layer lines are chiefly due to the oxygen atoms of the C=O groups, together with some effects due to small inequalities of the different zigzags in the repeating units. In considering the transition from the small pseudo-unit cell just mentioned to the true unit cell, it is necessary to take into account the different possibilities arising from the attachment of oxygen atoms to alternative positions on neighbouring chains; thus, in two dimensions, two possibilities are illustrated in figure 2. In three dimensions there are various possible unit cells, all compatible with the small pseudo-unit, but all could be ruled out except one, and this accounted well for all the spots on the photographs of the 66 polymer, except for a row of weak reflexions very near the meridian. (On most photographs, these spots are joined to other spots by streaks along the layer lines, and they may not be distinguished from such streaks; but suitably annealed specimens show them clearly—see figure 3L in Fuller *et al.* (1940); and pressed specimens, suitably oriented, show them more strongly and clearly—see our figure 9. The interpretation of the streaks will be considered later.)



a *b*

FIGURE 2

On considering the merits of a corresponding unit cell for the 6.10 polymer—a cell, that is, with the same angles but different height—it was again found that all the spots on all the layer lines fitted, except a row of weak spots near the meridian. This seemed highly significant, for individual spots on the 6.10 photographs are (except for those on the 9th layer) in quite different positions from those on the 66 photographs; but the interpretation could only be correct if it could be assumed that the weak spots near the meridian are due to a second crystalline phase present in the specimens of both polymers. This second crystalline phase would be simply a different packing of the same molecules (the repeat distance being exactly the same as in the main crystalline arrangement).

This is not the first occasion on which polymorphism of this type has been suggested to account for the X-ray diffraction patterns of long-chain polymers. Fuller & Frosch (1939) suggest that fibres of certain polyesters contain two different crystalline forms which can be derived from each other by simply displacing the chains parallel to each other (see figure 3). This is essentially the same as Schoon's suggestion (1938) to account for the various polymorphic forms of the shorter-chained hydrocarbons and acids (20 to 40 carbon atoms). The mode of side-by-side packing of the zigzag chains is the same if the displacements (*d* in figure 3) are exact multiples of the zigzag period 2·5 A, and displacements of this magnitude do account for the known unit cells. Various forms of cell—orthorhombic, monoclinic and triclinic—are obtainable by suitable displacements of the chains.

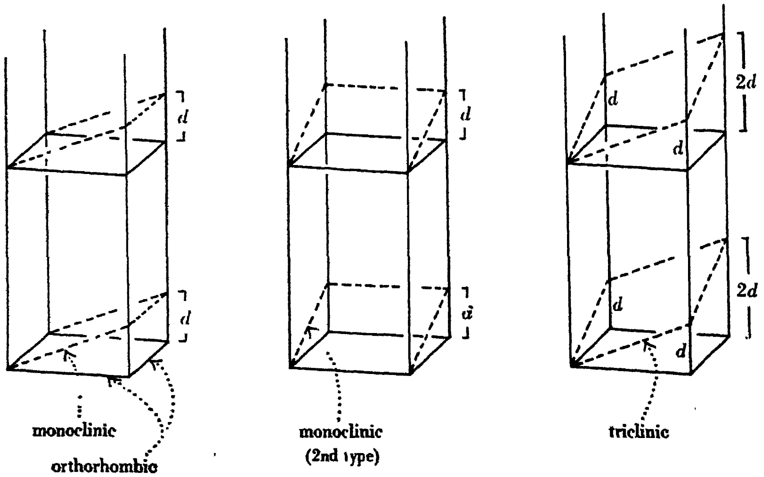


FIGURE 3. Polymorphism of chain compounds.

The hypothesis of polymorphism was accepted provisionally, and evidence in favour of it accumulated gradually, as the interpretation became more detailed. But we shall anticipate events by saying that, after the complete interpretation presented here had been worked out, a specimen of the 66 polymer was obtained which showed no trace of the spots ascribed to the second crystalline form (see figure 10, plate 1). The specimen was prepared by immersing a taut fibre (0·5 mm. diameter) for 10 min. in a 7 % solution of phenol in water, then for 2 hr. in boiling water, the phenol being removed by drying in air at 65° C for 3 days. This specimen evidently consisted of a single crystalline species; had it been available earlier, it would have simplified the process of interpretation. But its discovery strikingly vindicated the whole scheme of interpretation which had by that time been worked out.

The crystal form which is predominant in most nylon specimens will be called the α form. The dimensions of the triclinic unit cells of the α form of both polymers are shown in figure 4. The second crystal form, which gives the row of spots near the meridian of the fibre photograph, will be called the β form.

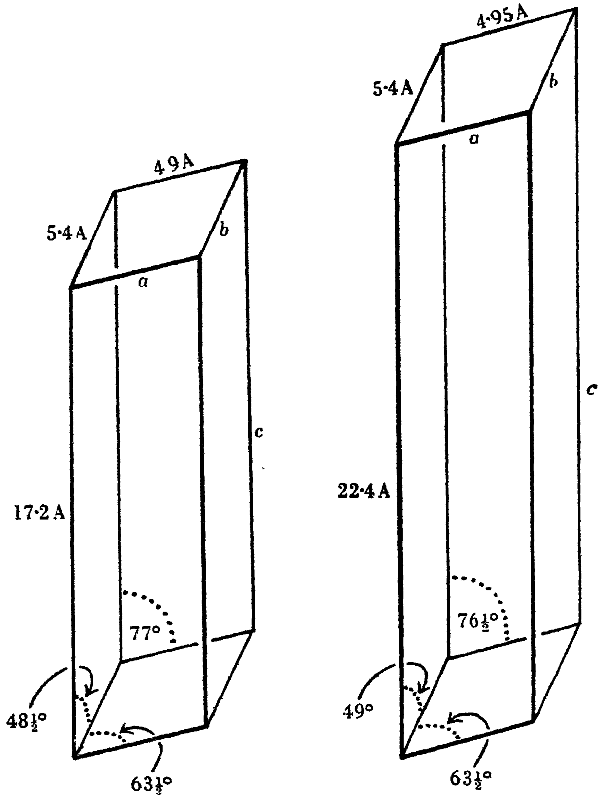


FIGURE 4. Dimensions of unit cells of 66 and 6.10 polyamides.

The next step was made possible by the discovery that a double orientation of the crystals can be obtained by pressing or rolling nylon fibres. In a fibre, all the crystals have their *c* axes parallel to each other, but in other respects are disposed at random; but in a pressed or rolled specimen, the crystals not only have their *c* axes parallel, but also have one particular crystal plane approximately parallel to the plane of the sheet. In a fully drawn fibre cold-rolled in the direction of the fibre axis, the *c* axes of the crystals remain parallel to the fibre axis, but if a long undrawn fibre is pressed, the *c* axes become perpendicular to the original fibre axis. The *c* axes set themselves along the direction of greatest flow during pressing. There is some departure from strict parallelism, especially for the second (plane) orientation. An X-ray photograph taken with the beam passing down the *c* axis (figure 7, plate 1) showed that the plane which has been called 010 of the unit cell is parallel to the plane of the sheet; the same photograph confirmed the angle between the 010 and 100 planes of the cell.

When a doubly-oriented specimen is set at any particular angle to the X-ray beam, the pattern produced is similar to that which would be given by a single crystal or a twinned crystal oscillated over a small angle, this angle corresponding to the dispersion of crystal orientation in the specimen. For the polyamides four different orientations of the unit cell in the specimen, all of them consistent with the restrictions mentioned (one axis common and one plane common), are possible (figure 5), and the specimen thus simulates, not a single crystal, but a quadruplet. A photograph of such a specimen placed at a particular angle to the X-ray beam shows certain reflexions only, and the intensities of them are not the same as in a fibre photograph (see figure 8, plate 1). By observing the specimen setting at which each reflexion is recorded at maximum intensity, it is possible to deduce the approximate orientations of all the reflecting planes with respect to each other and to the axes of the specimen. The graphical method which was used for interpreting these photographs is essentially Bernal's reciprocal lattice method (see Bernal 1926), suitably extended to deal with the fourfold twinning in the specimens.

All the reflexions appeared at the expected angles for both polymers. It must be emphasized that this test of the proposed unit cell is a severe one, and the result of it leaves very little doubt that the proposed triclinic unit cell for the α form is correct. From these photographs it was also possible to distinguish between alternative indices for certain spots. The photographs also indicated the orientations of the principal reflecting planes of the β form: the reflexions are strongest when the X-ray beam is parallel to the plane of the sheet, as in figure 9, plate 1.

Atomic positions in α crystals. The final test of any proposed polymer structure is to calculate the intensities of all the reflexions, the positions of the atoms in the cell being adjusted by trial in an attempt to obtain agreement between calculated and observed intensities. (Direct or semi-direct Fourier series methods of finding or refining atomic positions, which are used whenever possible for single-crystal photographs, are not practicable for most polymers.)

In the triclinic system, only two space-group symmetries are possible—$P1$, with no symmetry at all, and $P\bar{1}$, with a centre of symmetry. Molecules of 66 and 6.10 polyamides, by themselves, both have centres of symmetry, and assuming that this is utilized in the crystal arrangement (in this one-molecule cell, no way of avoiding its utilization is apparent), the space-group symmetry is $P\bar{1}$.

Ignoring hydrogens, whose positions cannot be found by X-ray methods, there are eight atoms to be placed in the cell—six carbons, a nitrogen and an oxygen. From the outset, there was no doubt about the approximate form of the molecule—a plane or nearly plane zigzag chain with oxygen atoms also lying in or near this plane; the problem of finding the atomic co-ordinates therefore resolves itself into the problem of finding the orientation of the molecule in the unit cell and detecting any distortions from the planar form.

It must be emphasized that the amount of information in most polymer photographs is not sufficient for a quite independent determination of all the atomic positions on X-ray intensity evidence alone, such as is possible when the very sharp, detailed photographs of single crystals are available. The limitations of polymer photographs are such that we must be content with accepting, for the most part, interatomic distances derived from other work, and moving and distorting the molecules without departing greatly from the accepted interatomic distances. This limitation in the interpretation of polymer photographs is not really a serious one,

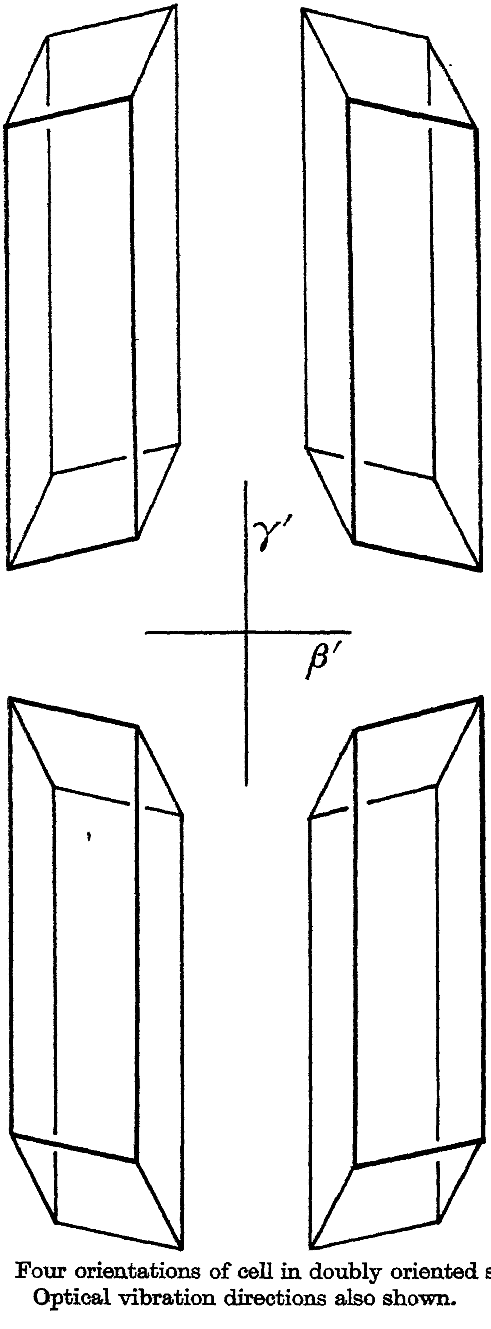
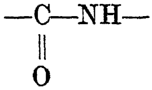
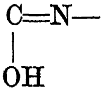


FIGURE 5. Four orientations of cell in doubly oriented specimens. Optical vibration directions also shown.

inasmuch as most interatomic distances, such as C—C, C—N, C═O, are well established and are not likely to vary appreciably; the only circumstances in which interatomic distances are in doubt are those in which bond character is in doubt. Thus, in the polyamides, the carbon-carbon links are undoubtedly single, and their length can be assumed confidently to be 1·53 A; but the group usually written  $-\underset{\underset{\displaystyle O}{\|}}{C}-NH-$ (amide form) may actually exist as  $\underset{\underset{\displaystyle OH}{|}}{C}=N-$ (imidol form). The carbon-oxygen and carbon-nitrogen distances are on this account uncertain; C—O is usually 1·43, C═O 1·25, C—N (in amides and amino-acids) 1·39, C═N 1·28. The procedure adopted was to keep the C—C bonds at 1·53 ± 0·03, but to treat the nitrogen and oxygen atoms semi-independently, moving them about in the cell regardless of interatomic distances within the wide limits between single- and double-bond distances. The diffracting powers of C, CH_2, NH and O were assumed to be in the ratios of the numbers of electrons 6 : 8 : 8 : 8, and the decrease of diffracting power with angle of diffraction to be as given for carbon by Cox & Goodwin (1936).

The approximate orientation of the molecule in the unit cell is suggested by the cell dimensions. The distance between chains A and D in figure 1 (4·77 A) is greater than that between A and B (4·03 A), and therefore the plane of the zigzag chain is expected to be in or near the (010) face of the cell. Furthermore, the angle β between the a and c axes is 77°, and this is just the angle to be expected if the molecules are lying with their planes in or near (010) and with the oxygen atom of one molecule on the same level as the NH of the next, as it would be if there is a hydrogen bond between them (see figure 2a).

Calculations of the relative intensities of the equatorial ($hk0$) reflexions and subsequently of the whole of the reflexions confirmed this, and defined the orientation more closely. In the calculations, the objective was to obtain relative intensities of the same order of magnitude as those observed on the photographs, the latter being estimated by eye and classified as strong, medium, weak and so on; for the interpretation of polymer photographs it is, generally speaking, scarcely justifiable to attempt more than this. The limitations are well brought out by experience in the present work, for it must be admitted that entirely satisfactory agreement between observed and calculated intensities, even within this modest scope, could not be attained; and it appeared that no adjustments of atomic coordinates could possibly remove all the discrepancies. This may mean, of course, that the whole scheme of interpretation is wrong; but against this it must be urged that, apart from the difficulty of finding the unit-cell dimensions (and the results here can now scarcely be doubted in view of the checks provided by the doubly-oriented specimens), the structure is a simple one, and in a one-molecule cell with only eight atoms to place and all of them closely limited in position by accepted interatomic distances and angles, the possibilities of adjustment are limited, and it can be said with more confidence than usual that the discrepancies cannot be removed by any adjustments of atomic positions. We shall return to these discrepancies later, and merely observe here that they are systematic, that they are of the same type in both 66 and 6.10 polymers, and that reasonable explanations have been found for them.

Accepting these systematic discrepancies, the best agreement between calculated and observed intensities was obtained by placing the plane of the chain nearly parallel to the 120 plane of the crystal, but with the oxygen atom pushed very nearly back to the 010 plane (see figure 12)—that is, appreciably out of the plane of the chain. In view of the uncertainties of interpretation, it may be doubted whether this feature is genuine; but in its favour it must be emphasized, first, that any other configuration gives a poorer agreement between calculated and observed intensities (involving non-systematic discrepancies), and secondly, that the determination of the chain orientation is largely independent of that of the position of the oxygen atom: the intensities of reflexions on the equator and on the 7th layer for 66 polymer (the 9th layer for 6.10) are determined mainly by the positions of the chain atoms, while the intensities of reflexions on all the other layer lines are determined mainly by the positions of the oxygen atoms.

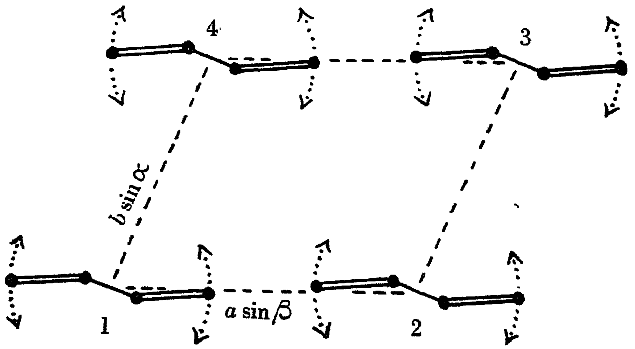


FIGURE 12. Structure of 66 and 6.10 polyamides. c projection. Arrows indicate movements of oxygen atoms when molecules rotate round their long axes.

Very similar results were obtained for the two polymers with which this work is concerned. Thus the structures are completely analogous; they are completely defined by the following figures. (The parameters given for the 66 polymer imply small distortions of the chain from the strictly planar form; but it is doubtful whether these are significant. The 6.10 parameters are those of a strictly planar chain.)

66 *polyamide. α crystals*

Unit cell	$a = 4{\cdot}9$ A	$\alpha = 48\frac{1}{2}°$	Space group $P\bar{1}$.
	$b = 5{\cdot}4$ A	$\beta = 77°$	
	$c = 17{\cdot}2$ A	$\gamma = 63\frac{1}{2}°$	

Atomic coordinates in fractions of unit cell edges:

CH_2 (1)	−0·083	CH_2 (2)	+0·120	C	−0·088	O	−0·371
	+0·041		−0·040		+0·010		+0·010
	+0·037		+0·115		+0·186		+0·201
N	+0·058	CH_2 (3)	−0·131	CH_2 (4)	+0·089	CH_2 (5)	−0·106
	−0·019		+0·095		−0·039		+0·065
	+0·253		+0·313		+0·392		+0·458

6.10 *polyamide.* α *crystals*

Unit cell
$a = 4{\cdot}95\,\mathrm{A}$ $\alpha = 49°$ Space group $P\bar{1}$.
$b = 5{\cdot}4\,\mathrm{A}$ $\beta = 76\frac{1}{2}°$
$c = 22{\cdot}4\,\mathrm{A}$ $\gamma = 63\frac{1}{2}°$

Atomic coordinates:

CH_2 (1)	+0·102	−0·046	+0·029
CH_2 (2)	−0·102	+0·046	+0·080
CH_2 (3)	+0·102	−0·046	+0·138
N	−0·083	+0·037	+0·187
C	+0·102	−0·046	+0·242
O	+0·358	−0·022	+0·228
CH_2 (4)	−0·102	+0·046	+0·296
CH_2 (5)	+0·102	−0·046	+0·358
CH_2 (6)	−0·102	+0·046	+0·411
CH_2 (7)	+0·102	−0·046	+0·473

The best values for the lengths of the c axes were obtained from photographs of specimens rotated about a direction perpendicular to the c axis (figure 18, plate 2). For each polymer a strong reflexion was produced from a plane almost exactly perpendicular to the fibre axis (1.3.14 for the 66 polymer, 1.3.18 for 6.10). The spacing of this plane is equal to the length of the c axis.

The complete structure of one unit cell of the 66 polymer is shown in figure 13, and a view along the c axis in figure 12.

Thermal vibrations and crystal distortions. The observed and calculated intensities are compared in an appendix, and a graphical comparison for the 66 polymer is shown in figure 14. It will be seen that there are some discrepancies. Taking the average correlation of calculated and observed intensities as standard, it is noticeable that for reflexions from the 010 plane and from all planes making a small angle with 010 (that is, all planes in which k has a high value in comparison with h, such as 020, 120, 122, 124, 126 in the case of 66 polymer), the calculated intensities are much larger than the observed, while the opposite is true for planes approximately perpendicular to 010 (such as 310, 212). No adjustments of atomic parameters can remove these discrepancies. The same is true for the 6.10 polymer. The fact that the discrepancies are for the most part systematic and of the same type for both polymers suggests that they are not due to incorrect parameters, but to some general feature of the structure.

Two possible explanations of these discrepancies have been considered, and it is likely that both play some part in producing the observed effects. One is that thermal vibrations of the molecules are strongly anisotropic. Vibrations of a triclinic structure are certain to be anisotropic, and in the case of the present proposed α-polyamide structure, in which the molecules are held together much more firmly in the 010 plane (by the hydrogen bonds—see below) than in other directions, it is likely that thermal vibrations will be greatest in the direction normal to the 010 plane and least *in* this plane; as a consequence, the intensities of reflexions from 010 and all planes making a small angle with 010 will be reduced in comparison with those from other planes. Many different types of thermal vibrations are, of course, possible in a structure of this type—sinuous movements of the chain molecules, torsional oscillations of the molecules about their axes, as well as vibrations of individual atoms. These different types of movement will affect the intensities in somewhat different ways, but in general the effects will be those already stated, which correspond with those actually observed.

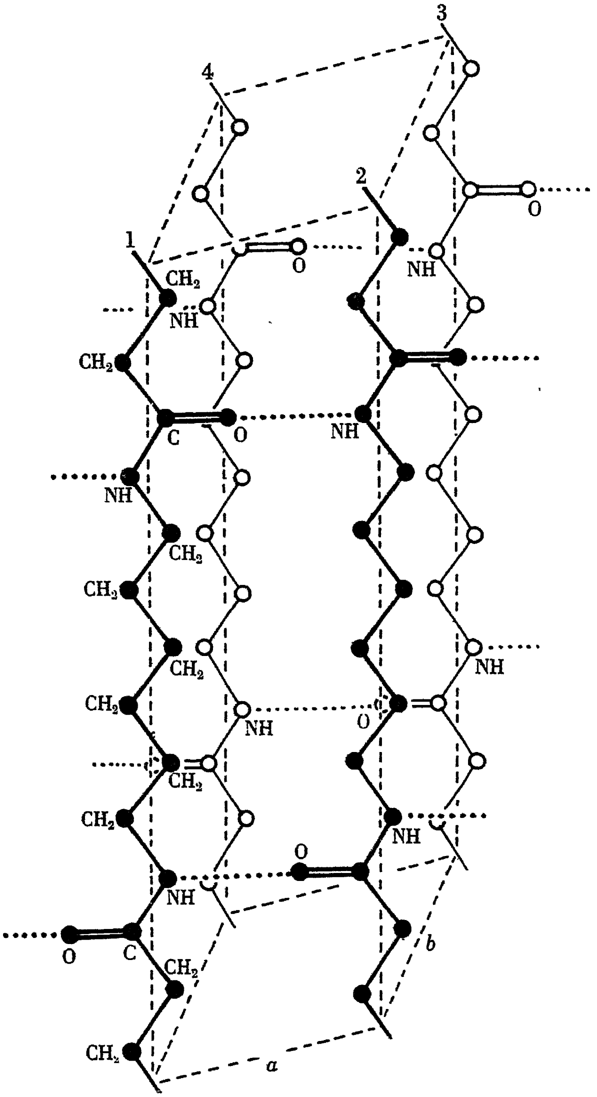


FIGURE 13. Structure of α crystal of 66 polyamide.

Torsional oscillations of the molecules about their axes may be responsible for the fact that the distances between carbon atoms in neighbouring molecules are not all the same. Those between atoms in molecules 2 and 4 in figure 12 (4·3 A) are perceptibly greater than those between atoms in molecules 1 and 2, or molecules 2 and

3 (3·9 to 4·0 A). The sweep of the oxygen atoms of molecule 2 would tend to push molecule 4 away, without affecting the distances between 1 and 2 or between 2 and 3.

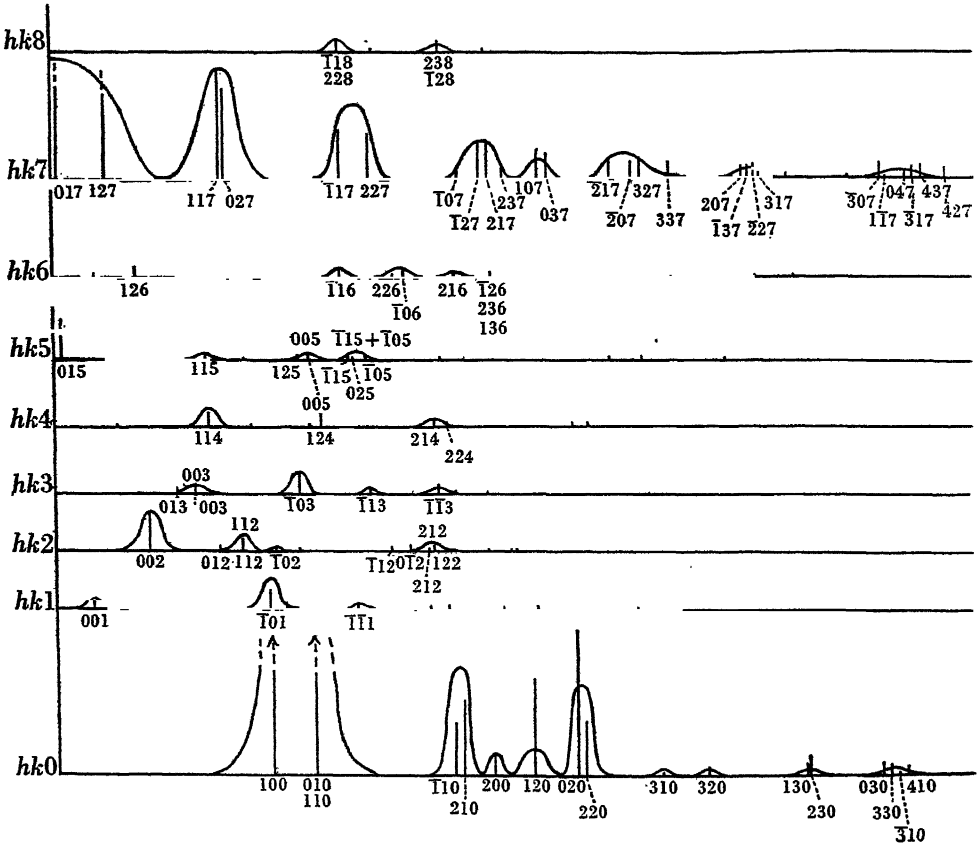


FIGURE 14. Comparison of observed and calculated intensities. 66 polyamide, α form. Observed $\overset{114}{\frown}$. Calculated $\underset{114}{\perp}$. Where photographs of doubly oriented specimens distinguish between alternative indices, verdict is shown thus $\overset{212}{\underset{212\ \bar{1}22}{\frown}}$.

The second possible explanation of the discrepancies in the intensities is that the crystals are likely to be more distorted in the direction normal to the *b* axis than in other directions. This again is a consequence of the layer formation of the crystal structure. Each layer of molecules would be flexible in directions perpendicular to the layer (figure 15), but molecules would be less easily distorted in the plane of the layer, owing to the much stronger forces between them. The effect on the intensities would be similar to that of the anisotropic thermal vibrations, and it does not seem possible to decide to what extent each of the effects is responsible for modification of the X-ray intensities.

Since the discrepancies can be explained in a reasonable way, it is considered that the structures proposed are acceptable interpretations of the X-ray data. In comparing the discrepancies with those encountered in an earlier case in which such effects have been discussed (Bunn 1939), it should be noted that the difference is that whereas in polyethylene all the $hk0$ reflexions are weakened to the same extent (because the lateral forces are much the same in all directions), in nylon the $0k0$ reflexions and those from nearby planes are weakened much more than $h00$ and nearby reflexions. Whether (and to what extent) $h00$ reflexions of nylon are weakened in comparison with those from planes normal to c—an effect which would be expected owing to the shape of the electron cloud of the CH_2 group (Bunn 1939)—we do not know. It is difficult to make a reliable comparison of these reflexions on a fibre photograph, and on a powder photograph there is too much overlapping of other reflexions.

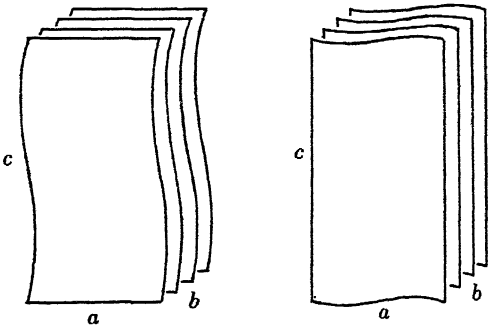


FIGURE 15. Distortion of crystals.

Structure of β form. The amount of information available for deducing the molecular arrangement in the β form is meagre; nevertheless, it has proved possible to draw certain definite conclusions about it.

The single set of spots available is a set of successive orders of reflexion by one particular plane, which is nearly at right angles to the molecular length (the c axis); the spots are evidently 001, 002, etc., of the β crystals. The spacings of these spots show that the molecular identity-period (the c axis) is indistinguishable from that of α crystals; the chain structure in β crystals is thus the same as in the α form. The difference is in the positions of the molecules along the c axis. The state of affairs is similar to that in certain polyesters (see figure 3), and contrasts with that in gutta-percha, which can be obtained in two different crystal forms consisting of geometrically different molecules (Bunn 1942).

The intensities of the spots on the set of photographs of doubly-oriented specimens make it possible to define approximately the orientation of the 001 plane. This orientation is such as to suggest a cell having an ac angle of about 77° (as in the α form) and a bc angle about 90° (contrasting with $48\frac{1}{2}°$ for the α form). The angle of 77° suggests that in the β form there are sheets of molecules joined by hydrogen bonds in the 010 plane, just as in the α form; the only difference between the two crystal forms would thus be in the relative positions of successive sheets of molecules. In the α form, if we take one particular sheet as the standard of reference, with the *c* axis vertical as usual, the next sheet is not on the same level as the first, but is displaced upwards (in our setting) by 3·55 A, this displacement defining the *bc* angle as 48½°. The third sheet is displaced in the same direction by the same amount, and so on (see figure 16). Since in the postulated β cell the *bc* angle is about 90°, it appears either that all the layers are on the same level, or else that there is a scheme of successive displacements, some up and some down, the net result being an eventual return to the same level. The simplest assumption is that all the layers are on the same level, in which case the unit cell would be a triclinic one-molecule cell. The relative intensities of the 00*l* reflexions are, however, such that this arrangement is out of the question; if all the molecules were on the same level, the relative 00*l* intensities would be practically the same for the β form as they are for the α form. In fact they are not; for 66 nylon, α 002 is very much stronger than α 001, whereas β 002 has about the same intensity as β 001 (see figure 9, plate 1).

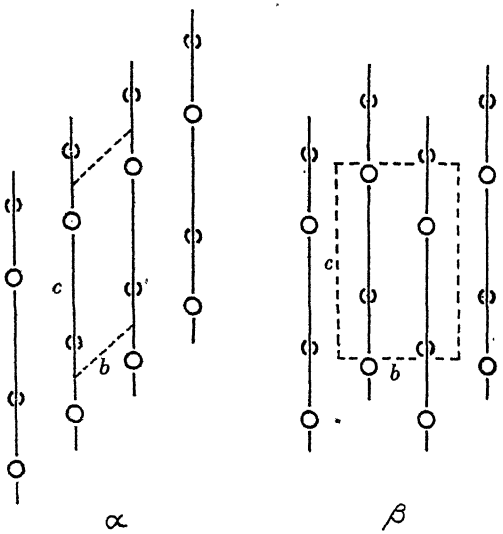


FIGURE 16. Arrangement of sheets of molecules in α and β crystals. Lines represent chain molecules, circles oxygen atoms.

The next possible structure (in order of simplicity) is one in which the second layer is displaced upwards with regard to the first, and the third is on the same level as the first (figure 16β). This arrangement has a triclinic two-molecule cell. It was found that such an arrangement does give correct relative intensities for the 00*l* reflexions, if the layer displacement is about 3·0 to 3·5 A (alternately up and down); moreover, the intensities of the β spots of the 6.10 polymer demand this same layer displacement (for figures, see Appendix). Note especially the figure for the layer displacement (3·0 to 3·5 A) which is about the same as in the α form (3·55 A); it appears, in fact, that the only difference between α and β crystals is that in the α form successive layers are displaced always in the same direction by a fixed distance 3·55 A, whereas in the β form successive layers are displaced alternately up and down by the same distance. This correspondence between the suggested α and β structures (the relation being the same in both 66 and 6.10 polyamides, and thoroughly in line with the polymorphism of other chain compounds) is one more significant point in favour of the whole scheme of interpretation of nylon photographs proposed in this paper.

The structures of β crystals of the two polymers are defined by the following figures. Only the coordinates of oxygen atoms are given; the amount of information available is insufficient for the determination of the coordinates of chain atoms. The structure of the chains (and, indeed, that of any one 010 layer of molecules) must, however, be very similar to that found in α crystals.

66 *polyamide. β crystals*

Unit cell $a = 4{\cdot}9$ A $\alpha = 90°$ Space group $P\bar{1}$.
$b = 8{\cdot}0$ A $\beta = 77°$
$c = 17{\cdot}2$ A $\gamma = 67°$

Chains lie along the lines $0{\cdot}0a$, $0{\cdot}25b$ and $0{\cdot}0a$, $-0{\cdot}25b$.
Oxygen atoms of reference molecule at $-0{\cdot}371a$, $-0{\cdot}250b$, $+0{\cdot}110c$
and $+0{\cdot}371a$, $-0{\cdot}250b$, $+0{\cdot}690c$

6.10 *polyamide. β crystals*

Unit cell $a = 4{\cdot}9$ A $\alpha = 90°$ Space group $P\bar{1}$.
$b = 8{\cdot}0$ A $\beta = 77°$
$c = 22{\cdot}4$ A $\gamma = 67°$

Chains lie along the lines $0{\cdot}0a$, $0{\cdot}25b$ and $0{\cdot}0a$, $-0{\cdot}25b$.
Oxygen atoms of reference molecule at $-0{\cdot}370a$, $-0{\cdot}250b$, $+0{\cdot}195c$
and $+0{\cdot}370a$, $-0{\cdot}250b$, $+0{\cdot}645c$

The reason why only one set of β reflexions appears on the photographs is that the proportion of β crystals in the specimens is small. The $00l$ reflexions appear because they fall very near the meridian of the photographs, where geometrical factors make for enhanced intensities (see Cox & Shaw 1930). Reflexions from $hk0$ and $hk7$ planes ($hk9$ for 6.10 polymer) should have appreciable intensities, but they necessarily coincide with reflexions from α crystals, on account of the close relation between the two arrangements.

Layer line streaks. All the spots on the X-ray photographs of 66 and 6.10 polyamides have now been accounted for. In addition to the spots, there is one unusual feature in the photographs of most specimens. On each of the lower layer lines of spots, near the meridian, is a streak connecting the spots. The most intense streak is on the second layer line, where it connects the strong α 002 spot with the weak β 002 spot. The intensities of the streaks vary in photographs of different specimens.

Such streaks, though unusual, have been seen before in X-ray photographs of several different types of substances—the fibrous mineral chrysotile (Warren & Bragg 1930), the fibrous protein keratin after stretching in steam (Astbury & Street 1931; Astbury & Woods 1933), and certain organic substances, such as *p*-azoxyanisol, which on heating are capable of forming a liquid crystal phase before complete melting occurs (Bernal & Crowfoot 1933). The explanation suggested by Warren & Bragg in the case of chrysotile is that one of the crystal axes—the *b* axis—is variable in length; the *a* and *c* axes are fixed by the nature of the structure, but it is possible to imagine alternative arrangements which would alter the *b* axis. For keratin, Astbury & Woods suggest a similar explanation, that one of the crystal axes is variable owing to a hydrolytic breakdown among the side-chain linkages. For the polyamides it is scarcely possible to entertain an explanation of this type, since the molecules are made from well-defined monomers of a high degree of purity, and the structure is such that variation of the side-spacings seems unlikely.

The test of such an explanation is whether or not streaks are formed on the equator as well as on the upper and lower layer lines. The position is usually confused by the presence of 'white' radiation in the X-ray beam, which produces equatorial streaks with quite normal crystals. A photograph of 66 polyamide was therefore taken with copper $K\alpha$ radiation monochromatized by reflexion from the 002 plane of a pentaerythritol crystal. On this photograph (figure 11, plate 1) streaks on upper and lower layer lines remain, but there are no streaks on the equator; hence there is no variation of the side-spacings.

The layer-line streaks found in X-ray photographs of substances like *p*-azoxyanisol have been interpreted by Bernal & Crowfoot (1933) as evidence that the longish molecules have a certain degree of freedom of position with respect to each other; they may slide along their major axes to a limited extent, thus causing the unit cell to be slightly variable in shape, this giving rise to reflexions drawn out along the layer lines. (At higher temperatures the molecules gain complete freedom of movement in some but not all respects, this type of phase being a typical liquid crystal; at still higher temperatures the substance becomes a true liquid in which all movements are possible.)

For the polyamides we suggest an explanation derived originally from the conception of two crystal forms differing from each other only in the relative positions of successive sheets of molecules. A large number of layers arranged as in figure 16 (left) constitutes an α crystal; a large number arranged as in figure 16 (right) constitutes a β crystal; but in some regions there may be parallel growths of β on α crystals (this is quite likely because the sheet structures are identical), and further, there may be some indiscriminate placing of layers as in figure 17, in which small sections of α and β groupings follow each other at random. Such a composite grouping would give an X-ray reflexion at a position intermediate between those of corresponding α and β reflexions, and the existence in the specimen of composite groupings containing all possible proportions of α and β types means that X-ray reflexions are produced at all positions between those of true α and β spots. The streaks must follow the layer lines, because whatever the mode of packing of the molecules, the repeat distance in the molecules (which determines the layer line spacing) remains constant. This is what is observed in the photographs—for instance, in figure 11, plate 1, α 002 is connected to β 002 by a continuous streak which follows the 2nd layer line (on photographs taken in a cylindrical camera it is a straight line). It is also worth noting that in figure 11, plate 1, the streak does not appear to cross the meridian of the photograph; it merely joins the α to the β spot without going beyond the β spot. This is consistent with the idea (embodied in the above explanation) that there is no disturbance of the structure of individual layers of molecules; the *ac* angle 77° in each layer is maintained, and this ensures that, whatever the positions of neighbouring layers of molecules (within the scheme given), the angle made by 002 with the *c* axis is never greater than 77°, and this is the reason why the streak does not cross the meridian of the photograph. (Only certain 66 photographs were sufficiently well defined to show this; in the 6.10 specimens the orientation of the crystals was not good enough.)

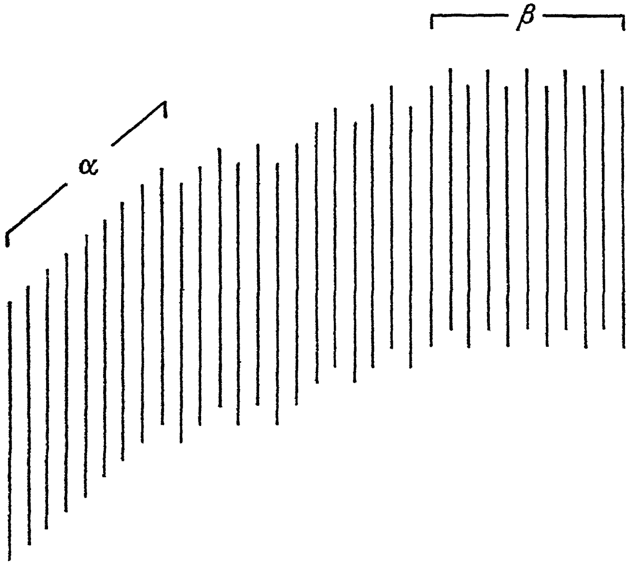


FIGURE 17. α, β and intermediate forms of nylon. Lines represent sheets of molecules seen edgewise.

This explanation is similar to that suggested by Bernal & Crowfoot for *p*-azoxyanisol, except that in the latter, continuous variation of molecular position over a limited range is postulated, whereas for the polyamides we suggest an indiscriminate succession of alternative, quite definite layer positions. It seems likely that in the polyamides the layer positions would be definite, to allow the chain zigzag to fit in the same way as in α and β crystals. There is some evidence in favour of this; fibres drawn at room temperature from quenched specimens (see Fuller *et al.* 1940), or drawn from the melt into cold air (see our figure 19, plate 2), give X-ray photographs showing no spots, but streaks only, on all except the equatorial layer. There is evidently much disorder in such fibres, but it is noteworthy that the streak on the 7th layer for 66 polymer (the 9th for 6.10) has a moderate intensity; the spacing for this layer is 2·45 A, the average zigzag repeat distance, and the intensity of the streak—comparable (in relation to the rest of the pattern) with that of the corresponding crystal reflexions of more ordered specimens—means that the zigzags are in register just as in the crystals. For a similar reason, the fact that the streak on the 2nd layer is strong means that the oxygen atoms are also in register as in the crystals. In fact, the relative intensities of the streaks on the different layer lines are consistent with our picture of composite α–β groupings rather than with a more complete disorder.

The fact that it is possible to explain the layer-line streaks in a manner arising naturally from the two-crystal theory is regarded as another point in favour of the present scheme of interpretation. The complexity of the interpretation (two triclinic crystal forms together with composite groupings) is matched by the wealth of evidence (positions and intensities of spots, orientation of crystal planes, and the intensities of layer-line streaks).

Amorphous material. The only remaining feature of polyamide fibre photographs is the diffuse ring passing through the first intense equatorial spot. Most X-ray photographs of polymers show such a ring, which indicates the presence of a certain proportion of molecules or sections of molecules not arranged or oriented in any precise way. The distances between such molecules are not constant but are likely to be equal or a little greater than the side-spacings in the crystals, and this is why the diameter of the ring is such that it passes through the inner of the two intense equatorial spots which show the side-spacings in the crystals. (The inner of the two rings corresponds to the greater spacing.)

The possibility of the presence in the amorphous material of molecules not having the straight-chain structure of the bulk of the material must not be overlooked. But observe in the first place that in view of the spacing of the amorphous ring (4·5 A) there is no room for side-chains. Secondly, the presence of significant amounts of large ring molecules (which would assume a long loop form with side-spacing similar to those of the chains—see Müller (1933)—is considered unlikely by Carothers (1940)).

All the features of X-ray diffraction photographs of 66 and 6.10 polyamides have now been accounted for. One is now in a position to consider the intermolecular forces in the crystal, and then to pass on to questions of texture and physical properties.

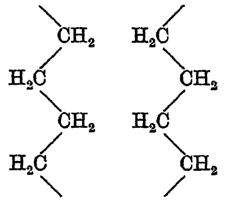
Interatomic distances and intermolecular forces. The most striking feature of the α structure (figure 13) is that the oxygen of one molecule is found opposite the nitrogen of the next; and the distance between these two atoms is 2·8 A, in both the 66 and 6.10 polymers. The Van der Waals radius of NH is about 1·8 A and that of O 1·35 A (see Albrecht & Corey 1939), hence if there were no electrostatic interaction the N—O distance would be expected to be $1\cdot8 + 1\cdot35 = 3\cdot15$ A. The shorter distance found, 2·8 A, is characteristic of hydrogen bond formation: it is in good agreement with the N—O distances found in glycine (Albrecht & Corey 1939), acetamide (Senti & Harker 1940), alanine (Levy & Corey 1941) and diketopiperazine (Corey 1938).

The question whether the molecules have the amide or imidol structure is not answered by the present work: atomic parameters cannot be determined with sufficient accuracy. This is illustrated by the figures for the C—O distance, which in the 66 polymer appears to be about 1·35 A and in the 6.10 polymer 1·28 A; it is obviously not possible to decide whether the bond is single (for which the usual distance is 1·43 A) or double (1·25 A). The distance from carbonyl carbon to nitrogen appears to be about the same in both polymers ($\sim$ 1·40 A); but this must be regarded as fortuitous, although the figure does agree with that found in other amides.

Apart from the hydrogen bond distance between N and O atoms, all the other distances between atoms of neighbouring molecules are the normal ones characteristic of the Van der Waals forces. The hydrogen bonds link molecules together in sheets in the 010 plane; but the links in other directions are much weaker, and the whole structure is thus of a pronouncedly layer-like type. It is this feature which is responsible for the plane orientation in rolled or pressed sheets—the sheets of molecules linked together by hydrogen bonds are found parallel to the plane of the sheet of material. In a more general way, it is likely that the high tensile strength of nylon fibres, which enhances their practical value, owes something to the hydrogen bonds: when a fibre is broken it is likely that individual chains are not broken, but that crystals are pulled apart by shear (molecules slipping past each other parallel to their lengths), and this would be the more difficult, the stronger the forces holding the molecules side by side.

The other features of the crystal structure we have to try to understand are: (1) the orientation of the plane of the zigzag about 20° to the 010 plane, and (2) the mode of packing of the layers.

(1) The relative positions of any two neighbouring molecules in one layer (e.g. 1 and 2 in figure 13) are settled primarily by the hydrogen bonds, which keep the oxygen atom of one molecule on the same level as the nitrogen of the next, and fix the *ac* angle of the cell at 77°. But this has the effect of placing the CH_2 groups awkwardly:



These groups do not interlock; the projections of one chain are on the same level as the projections of the next. This is presumably the reason (or one of the reasons) why the zigzag chains turn away from the 010 plane; it is an attempt to relieve the packing situation. Possibly the positions of the hydrogen nuclei have something to do with the actual angle assumed. The principles of packing of hydrocarbon molecules are not entirely clear, but it may be that the CH_2 group should not be regarded as an unresolved packing unit (as the electron density distribution might suggest), but that the positions of hydrogen nuclei are important. Mack (1932) has suggested that in several hydrocarbons the packing is determined by this sort of interlocking of hydrogens, while the arrangement in polyethylene (Bunn 1939) may be regarded in the same way. Examination of a model of the present polyamide structure suggests that here also the arrangement of hydrogen nuclei may play a part in determining the orientation of the chains.

(2) As mentioned above, any one layer is displaced along the *c* axis by 3·55 A with regard to its neighbouring layer. Why the distance 3·55 A?

In the first place, the layers would not be expected to be on the same level, because such an arrangement would not interlock; this is true, whether a CH_2 group is regarded as an unresolved packing unit or whether one takes into account Mack's principle of hydrogen interlocking. The side-by-side positions of the layers are, of course, also concerned in the packing; but here there is no alternative to the arrangement in figure 12, where molecule 4 fits into the hollow between 1 and 2. The available levels consistent with good packing* are those with displacements of about

$$\frac{c}{14}, \frac{3c}{14}, \frac{5c}{14}\ (= 1{\cdot}23,\ 3{\cdot}68,\ 6{\cdot}14\,\mathrm{A})$$

and so on. Obviously the first-mentioned position would not be favoured, because the polarized groups would be too close: arrangements of this sort

$$\overset{+}{\mathrm{H}}\text{—N—C}=\overset{-}{\mathrm{O}},\quad \overset{+}{\mathrm{H}}\text{—N—C}=\overset{-}{\mathrm{O}}$$

are not likely to be stable. In the second position, with a displacement of about 3·5 A (that actually found in the crystals), these polarized groups would be sufficiently far apart to have little interaction, but it is difficult to see why this second position should be any better than the third (to say nothing of other positions); any explanation must be valid for both 66 and 6.10 polymers, and so far no satisfactory explanation has been found. Presumably this will become clearer as knowledge of similar structures accumulates.

Optical properties. In the triclinic system, the three principal vibration directions—the three axes of the indicatrix—are not necessarily related in any simple way to the axes of the unit cell. To measure the three principal refractive indices associated with these vibration directions, it would be necessary to have single crystals. These are not available; the nearest approach is a doubly-oriented strip in which there are four different orientations of the unit cell. However, the structure is such that the refractive indices for the vibration directions defined in the doubly-oriented strip are likely to be near the true principal values, since the principal axes of the indicatrix are likely to lie near the axes of the doubly-oriented strip. The three indices of the strip will be called α', β' and γ'; α' is the index for light vibrating perpendicular to the plane of the strip (perpendicular to the molecular sheets), β' the index for light vibrating in the plane of the strip and at right angles to the molecular axes (that is, along the C=O bonds), and γ' the index for light vibrating along the molecules (see figure 5).

For the 6.10 polymer, β' and γ' were measured on the Abbé refractometer; the strip, suitably oriented, was put in optical contact with the prism by a film of bromonaphthalene. The remaining index α' was obtained by making use of the properties of a fibre, which gives γ' for light vibrating along the fibre and the average of α' and β' for light vibrating across the fibre (on account of the random orientation of the crystals in this plane). From the value of $\frac{1}{2}(\alpha'+\beta')$ thus measured (1·500—this was done under the microscope by the immersion method), α' was calculated. The three indices are $\alpha' = 1\cdot475$, $\beta' = 1\cdot525$, $\gamma' = 1\cdot565$.

---

* It is impossible to illustrate these points by diagrams. The remarks made here are based on a detailed examination of a scale model.

For the 66 polymer, $\gamma'(= 1\cdot580)$ was obtained from a doubly-oriented strip on the Abbé refractometer, and also from a fibre under the microscope. $\frac{1}{2}(\alpha'+\beta')\,(= 1\cdot520)$ was also obtained from the fibre (as for the 6.10 polymer). To calculate α' and β' separately, use was made of the interference figure produced in convergent polarized light. A doubly-oriented strip viewed normally gives a figure very similar to that of a biaxial single crystal, except that there are no dark hyperbolae passing through the 'eyes' (figure 20*b*, plate 2). Ignoring for a moment this abnormality, assume that the distance between the 'eyes' is a measure of the optic axial angle. By comparison with crystals of known optic axial angle, it was found that the angle $2V$ for the 66 polymer is about $-45°$. From the approximate relation $\cos^2 V = (\beta'-\alpha')/(\gamma'-\alpha')$, and from the already known values of γ' and $\frac{1}{2}(\alpha'+\beta')$, the two indices α' and β' were calculated. The three indices are $\alpha' = 1\cdot475$, $\beta' = 1\cdot565$, $\gamma' = 1\cdot580$.

The abnormality of the interference figure—the absence of dark hyperbolae through the 'eyes'—is probably due to a variation of the optic axial angle in different small regions of the field of view or in layers at different distances from the surface of the sheet. Such a variation of the optic axial angle would be expected owing to the occurrence of different degrees of perfection of orientation of the crystallites; this would not affect the lemniscates to any great extent, but would remove the hyperbolae. If this explanation is correct, the positions of the 'eyes' would give a reliable though approximate indication of the average optic axial angle.

Some doubly-oriented sheets showed, in some regions, much more abnormal effects; one example is illustrated in figure 20*c*, plate 2. It is possible that such effects are due to the presence of a small proportion of crystals having a different orientation from the majority, depending on the direction of flow in different parts of the specimen during pressing or rolling. Experiments with mica sheets showed that such effects can be simulated by superimposing a very thin sheet on a thicker sheet, the optic axial planes being at right angles. It requires very little of a secondary reversed orientation to produce these effects.

The interference figure given by 6.10 specimens (figure 20*a*, plate 2) shows that the optic axial angle is greater than that of 66; this is in accordance with the measured values of the refractive indices, which give $2V = 90°$.

The maximum refractive index is, as would be expected, for light vibrating along the molecules; this is a consequence of their chain character. Of the other two indices, the greater is for light vibrating in the plane of the strip, that is, along the C=O bonds; this is in accordance with expectation, since double-bonded groups have a higher index for light vibrating along the bond than for light vibrating across the bond. And it is noteworthy that β' is higher for the 66 than for the 6.10 polymer;

this is due to the higher proportion of C=O groups in the former. Finally, α' is the same for both polymers; this is not surprising, since the vibration direction for α' is across the double bonds, which therefore would not affect the value of the index.

Thus the optical properties of polyamide specimens are thoroughly consistent with the structure deduced by X-ray methods. (The optical properties of the β crystals in the specimens would be expected to be very similar to those of α crystals.)

Texture of nylon specimens. Unoriented nylon is opaque, except in very thin sheets; in this it resembles polyethylene, guttapercha and other crystalline polymers. The optical discontinuities responsible for the opacity are not boundaries of single crystals, for the crystals are too small to be seen microscopically and, judging from the diffuse character of the X-ray reflexions, are probably much smaller than 1000 A in diameter—that is, very much smaller than the wave-length of visible light. The optical discontinuities are the boundaries of spherulitic aggregates of crystals. The spherulitic structure can be seen clearly in thin films: when examined between crossed Nicols, the black crosses indicative of a radial arrangement can be seen. The polyamides thus form additional examples of high polymers having a spherulitic structure, and this strengthens the impression (see Bunn & Alcock 1945) that this may be the normal manner of growth of crystalline high polymers generally.

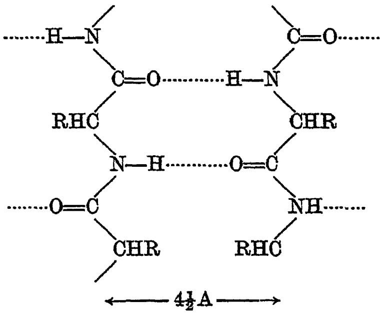
In spherulites of polyethylene and rubber, the optical properties show that the molecules are tangentially disposed; and this is consistent with the view that crystals grow outward from a point by side-by-side accretion of portions of molecules. In nylon spherulites, the use of the quartz wedge shows that for the 66 polymer the refractive index for the radial vibration direction is higher than for the tangential vibration direction; this is opposite to the state of affairs in polyethylene and rubber, and might seem to suggest that in nylon 66 the molecules are radially disposed. But this is not necessarily so; the optical properties are equally consistent with the view that the crystal direction which lies along the radius of a spherulite is the a axis or the nearby hydrogen bond direction. The refractive index for this direction (β') is, for the 66 polymer, nearly as high as γ', and a spherulite having this structure would have β' as its radial refractive index and $\frac{1}{2}(\alpha'+\gamma')$ (which is less than β') as its tangential refractive index. (Along any one radius there are crystals having all orientations obtainable by rotating about the β' direction.) Thus the optical properties are consistent with the radial direction being either the c axis (the molecular axis) or the hydrogen bond direction. The latter seems more likely to be correct, since crystal growth in polymers seems to be a matter of side-by-side accretion of portions of molecules; and one would expect the formation of sheets by growth along the hydrogen bond directions.

It would be expected that individual crystalline regions would be larger along the a axis than along the b axis. In X-ray diffraction photographs of fibres there is some evidence which is consistent with this: reflexions from planes like 010, 120 are rather more diffuse than those from 100; this is consistent with a smaller dimension along b than along a, though it does not prove it, since there are other possible causes of the broadening of reflexions.

When unoriented specimens are drawn into fibres, the crystals become parallel to the fibre axis. Fankuchen & Mark (1944) have taken X-ray photographs at various points on the shoulder between drawn and undrawn portions of a fibre, and have found evidence that the first thing that happens is that a particular plane becomes oriented parallel to the fibre axis; later, this 'ribbon' orientation is succeeded by the final chain orientation. The evidence is that, of the two strong rings which are the outstanding features of X-ray photographs of unoriented material, the outer one becomes contracted to equatorial arcs at an earlier stage than the inner one. The present work has shown that the outer one consists of reflexions from 010 and 110 planes, the former being the stronger, while the inner one is 100. Evidently the 010 plane becomes oriented first; this is natural enough, since this is the plane in which molecules are held together strongly by hydrogen bonds. A specimen may be regarded crudely as a number of flat plates tied together by strings; on drawing, the first thing that happens is that the plates tend to become parallel to the direction of drawing.

Crystallographic relations between polyamides and the fibrous proteins. The crystal structures of the fibrous proteins such as silk fibroin and keratin (hair) are not known in detail, but some of the principal spacings and other features are known, and the relations between these and the corresponding features in polyamide crystals are worth noting.

In the first place, the fibrous proteins have a side-spacing of about 4·5 A which is remarkably constant from one protein to another (Astbury 1940); this is what Astbury calls the 'backbone spacing'—the distance which shows the spacing of the main chains in one particular crystal direction. The other main side-spacing, in a direction roughly at right angles to the backbone-spacing, varies from one protein to another (in β-keratin, for instance, it is 9·8 A, in silk fibroin 4·3 A), and is apparently due to the side-chains, the size of which varies in different proteins. It is considered by some writers that the constancy of the spacing of 4·5 A is an indication that the chains are packed in a similar way in the various proteins, and that hydrogen bonds are responsible for the type of packing:



The two side-spacings in 66 and 6.10 polyamide crystals are 4·4 and 3·7 A. The former of these, 4·4 A, is the spacing of the 100 planes of the crystals and it has been shown that the 100 planes are bridged by CO·······HN hydrogen bonds. And this is not all: in doubly-oriented strips of polyamides made by pressing or rolling, it has been found that the sheets of molecules joined by hydrogen bonds lie *in* the plane of the strip. Astbury & Sisson (1935) have made doubly-oriented keratin strips by compressing pieces of horn in an atmosphere of steam, and find that the crystal plane having the spacing 4·5 A is normal to the plane of the strip; in other words, if hydrogen bonds bridge the 4·5 A plane, the hydrogen bonds lie *in* the plane of the strip. (The side-chains are normal to the plane of the strip.) The correspondence is complete.

The orientation of crystals in pressed sheets is presumably determined by the relative dimensions of the crystals and on their cleavages. Polyamide crystals are probably larger along the *a* axis than at right angles to it, and on pressing a specimen, the larger dimension of a crystal would be expected to set itself in the plane of the sheet; in addition, cleavage would occur most readily along 010, making the crystals still thinner in the direction normal to the *a* axis, and thus the tendency for the *a* axis to lie in the plane of the sheet would be increased. Keratin crystals are likewise probably larger in the direction thought to contain the hydrogen bonds than in the other lateral direction (in this case the side-chain direction) (see Astbury & Sisson 1935); cleavage is less likely to occur in keratin, because some of the side-chains are thought to join main-chains by primary bonds.

Another correspondence is revealed when the whole X-ray pattern of keratin is compared with those of the polyamides. Astbury & Street (1931) remark that all but two of the spots in the β-keratin photograph belong to a single zone—the zone of planes perpendicular to the plane in which the hydrogen bonds are thought to lie. The photographs of polyamides are similar in this respect; planes perpendicular to the hydrogen bond plane give relatively strong reflexions, while those making small angles with the hydrogen bond give abnormally weak reflexions. The explanation suggested by Astbury & Street for the non-appearance of the theoretical number of reflexions in the case of keratin is that it is due to imperfect or variable intermolecular junctions in the side-chain direction—variable owing to the varying sizes of side-groups at different points along the main chains. This does seem a reasonable explanation; but the fact that similar features are shown by the polyamides, in which there are no side-chains, suggests that there may be an alternative explanation not involving side-groups; and we suggest that the explanation we have put forward for the polyamides—anisotropic thermal vibrations and crystal distortions—may also account, at any rate partly, for the phenomenon as it occurs in β-keratin.

Yet another correspondence is found in the layer-line streaks which occur on X-ray photographs of both steam-stretched keratin and the polyamides. In the case of the polyamides, such streaks, since they occur only on upper and lower layer lines and not on the equator of the fibre photograph, cannot be due to variation of side-spacings, and we have suggested that they are due to indiscriminate displacements of successive layers of molecules along the fibre axis. In the case of steam-stretched keratin, Astbury & Woods (1933) suggest that the streaks, which are here found on the equator also, are due to the effects of a hydrolytic breakdown among the side-chain linkages. Now this hydrolytic breakdown may have two different crystallographic effects; one is to modify the side-spacings, the other is to break the covalent disulphide cross-links, and thus to make possible some slipping of hydrogen-bonded layers past one another. This last effect would give rise to the same sort of partial disorder as that which apparently exists in nylon.

All these very striking correspondences between the X-ray patterns of the polyamides and the fibrous proteins suggest that the crystals are built on the same general plan. This plan is now known for two polyamides, and it is encouraging to note that the current ideas on the interpretation of protein X-ray patterns do correspond in a general way with what is now known of the polyamides: the conception of protein molecules joined in sheets by hydrogen bonds corresponds with the known structure of the polyamides. The expectation of finding some general structural features common to the polyamides and the fibrous proteins has thus been justified; and, moreover, the interpretation of certain details in the X-ray patterns of the polyamides has led to suggestions of corresponding interpretations for the fibrous proteins.

APPENDIX

Intensities of X-ray reflexions

The three columns give: indices, calculated intensity, observed intensity. The figures given for calculated intensities are proportional to the square of the structure amplitude F multiplied by Lorentz and polarization factors, and by a temperature factor $\exp\left[\frac{-B\sin^2\theta}{\lambda^2}\right]$, where $B = 5$, this being the figure found appropriate for the $hk0$ reflexions of polyethylene (Bunn 1939).

66 polyamide. α form

100	762	vs	001	9	w	002	45	m	013	5	—	014	2	—
010	1212	vvs	011	0	m	012	6	—	003	10	vw	114	18	w
110	704		$\bar{1}01$	25		112	15	w	113	1	—	004	2	—
$\bar{1}10$	64	s	111	1		$\bar{1}02$	6	—	$\bar{1}03$	27	w	$\bar{1}04$	3	—
210	90		101	6	—	102	2	—	103	0	—	124	13	—
200	23	w	$\bar{1}\bar{1}1$	9	vvw	$\bar{1}12$	4	—	$\bar{1}13$	9	vvw	$\bar{1}14$	1	—
120	119	w	$0\bar{1}1$	1		$\bar{1}\bar{1}2$	1	—	123	0	—	104	1	—
020	179	ms	$\bar{1}11$	5	—	$0\bar{1}2$	8	—	213	1	—	024	0	—
220	64		211	5	—	212	4	vw	023	2	—	214	9	vw
310	2	vvw	121	0	—	122	7		$\bar{1}\bar{1}3$	10	vvw	224	5	—
$\bar{2}10$	0		$\bar{2}01$	0	—	022	0	—	$0\bar{1}3$	2	—	$\bar{1}\bar{1}4$	1	—
$\bar{1}20$	2	vvw	$\bar{2}\bar{1}1$	1	—	$\bar{2}02$	2	—	223	2	—	$\bar{2}04$	1	—
320	9		$1\bar{1}1$	2	—	222	1	—	$\bar{2}03$	0	—	$0\bar{1}4$	1	—
300	1		201	1	—	$\bar{2}\bar{1}2$	2	—	203	0	—	$\bar{1}24$	0	—
130	14	vvw	021	0	—	202	2	—	$\bar{2}\bar{1}3$	0	—	204	4	—
230	24		221	4	—	$1\bar{1}2$	0	—	$1\bar{1}3$	0	—	$\bar{2}14$	0	—
030	14	—	$\bar{1}\bar{2}1$	1	—	$\bar{2}12$	0	—	$\bar{1}23$	0	—	134	4	—
330	11	vvw	$0\bar{2}1$	0	—	$1\bar{2}2$	0	—	$\bar{2}13$	0	—	$\bar{2}\bar{1}4$	0	—
$\bar{3}10$	2		$\bar{2}11$	2	—	$\bar{1}\bar{2}2$	0	—	133	0	vvw	234	0	—
410	9		$\bar{2}\bar{2}1$	0	—	312	0	—	323	0		324	1	—
			311	1	—	322	0	—	313	1		$1\bar{1}4$	0	—
			$\bar{1}21$	0	—	$\bar{2}\bar{2}2$	1	—	233	1		314	1	—
			$\bar{3}\bar{1}1$	0	—	$0\bar{2}2$	1	—				034	0	—
						132	2	—						

66 *polyamide. α form* (*cont.*)

015	large*	m	116	3	—	017	large*	s	1̄1̄7	1	vw	118	0	—
115	2	vvw	126	11	—	127	large*		01̄7	2		228	2	w
125	4	—	026	0	—	117	131	s	207	14		1̄18	8	
005	7	vvw	1̄16	7	vw	027	105		1̄37	15		138	2	—
1̄15	6	vw	006	0	—	1̄17	60	ms	247	1		238	2	vw
025	3		226	3	vw	227	55		2̄27	17		1̄28	9	
1̄05	6		1̄06	5		007	0	—	317	15		008	1	—
225	0	—	216	4	vvw?	137	0	—	147	1		038	1	—
105	0	—	106	0	—	1̄07	13	m	2̄1̄7	7	—	1̄08	2	—
215	3	—	1̄26	3	—	1̄27	43		347	4	—	218	0	—
1̄25	1	—	136	2		217	41		3̄07	19	vw	108	0	—
2̄05	0	—	236	1		237	12		11̄7	6		2̄18	0	—
135	1	—	2̄06	0	—	107	35	w	047	12		328	1	—
1̄1̄5	1	—	036	0	—	037	28		3̄17	15		338	1	—
235	0	—	2̄16	0	—	2̄17	26	mw	437	17				
01̄5	3	—	326	0	—	2̄07	21		427	13	—			
2̄15	0	—	1̄1̄6	2	—	327	26							
205	3	—	01̄6	0	—	337	20	—						
325	4	—	206	2	—									
035	1	—	336	1										
2̄1̄5	0	—												
315	4	—												

* Reflexions on meridian; cannot be calculated precisely.

6.10 *polyamide. α form*

100	1330	vs	001	17	mw	002	49	s	003	10	vw	014	10	m
010	1850	vvs	1̄01	7	mw	012	5	vvw	013	7	w	114	11	
110	1240		101	4		112	1	—	113	13		1̄04	14	w
210	90	s	011	1		1̄02	14	w	1̄03	3	—	104	0	—
1̄10	92		111	6		102	2	—	103	4	—	1̄14	7	vvw
200	33	m	1̄1̄1	6	vvw	1̄1̄2	2	—	11̄3	1	—	124	0	—
120	166	m	01̄1	0	—	01̄2	4	vvw				214	1	—
020	106	ms	1̄11	0	—	1̄12	3	vvw	013	1	vvw	1̄1̄4	3	—
220	59		211	5	vvw	212	1	—	213	4		024	0	
310	1	vw	2̄1̄1	2	—	122	0	—	123	2		01̄4	1	—
2̄10	0		1̄11	0	—	2̄02	0	—	023	1	—	224	1	
1̄20	3	vw	2̄01	2		2̄1̄2	0	—	2̄03	2	—			
300	2		121	1		022	1	—	223	1				
320	3		201	2	—	202	0		2̄1̄3	1	—			
230	19	vvw	021	0	—	11̄2	0		203	2	—			
130	22		221	1	—	222	1							
030	20	vvw												
330	8													
3̄10	15													
410	17													

015	16	?*	016	5	?	117	4	—	118	6	—
115	11	w	116	1	—	127	2	—	12̄8	4	—
005	7		006	2	—	007	3		028	2	—
1̄05	0	—	126	1	—	1̄17	3	—	1̄18	24	w
125	3	—	1̄06	9	w	027	1	—	008	1	—
1̄15	0		1̄16	8		1̄07	1	—	1̄08	7	—
025	0		026	0	—	227	1	—	228	1	—
215	3	—	106	0	—	107	1	—	218	0	—
225	1		226	2	—				108	4	—
1̄1̄5	0	—	216	1	—						

* Obscured by diffuse ring due to amorphous material.

6.10 *polyamide. α form* (*cont.*)

129	52	?	$\bar{2}09$	24	vvw	0.2.10	0	—
119	145	s	339	18		1.1.10	9	—
029	74		$\bar{1}\bar{1}9$	2	vw	2.2.10	0	—
$\bar{1}19$	75	ms	$0\bar{1}9$	3		$\bar{1}$.1.10	16	vvw
229	56		209	28		0.0.10	0	—
009	0	—	319	24		$\bar{1}$.0.10	2	—
$\bar{1}09$	18	ms	$\bar{1}39$	24		2.1.10	0	—
$\bar{1}29$	46		249	3				
239	38		149	2	—			
219	59		$\bar{2}\bar{1}9$	10	vvw			
139	0	—	349	9				
109	38	vvw	$1\bar{1}9$	19	vw			
$\bar{2}19$	30	vw	$\bar{3}09$	24				
329	31		429	20				
			439	24				
			049	18				

66 *polyamide. β form*

001	33	s
002	28	s
003	8	w
004	25	m?*
005	53	s
006	0	—

6.10 *polyamide. β form*

001	0	—
002	49	s
003	1	—
004	7	w
005	12	?*
006	8	m
007	10	m
008	0	—

* Obscured by diffuse ring due to amorphous material.

REFERENCES

Albrecht & Corey 1939 *J. Amer. Chem. Soc.* **61**, 1087.
Astbury 1940 *Trans. Faraday Soc.* **36**, 871.
Astbury & Sisson 1935 *Proc. Roy. Soc.* A, **150**, 533.
Astbury & Street 1931 *Phil. Trans.* A, **230**, 75.
Astbury & Woods 1933 *Phil. Trans.* A, **232**, 333.
Bernal 1926 *Proc. Roy. Soc.* A, **113**, 117.
Bernal & Crowfoot 1933 *Trans. Faraday Soc.* **29**, 1032.
Bunn 1939 *Trans. Faraday Soc.* **35**, 482.
Bunn 1942 *Proc. Roy. Soc.* A, **180**, 40.
Bunn & Alcock 1945 *Trans. Faraday Soc.* **41**, 317.
Carothers 1940 *Collected papers on high polymeric substances.* New York: Interscience Publishers.
Corey 1938 *J. Amer. Chem. Soc.* **60**, 1598.
Cox & Goodwin 1936 *J. Chem. Soc.* p. 769.
Cox & Shaw 1930 *Proc. Roy. Soc.* A, **127**, 71.
Fankuchen & Mark 1944 *J. Appl. Phys.* **15**, 364.
Fuller 1940 *Chem. Rev.* **26**, 143.
Fuller, Baker & Pape 1940 *J. Amer. Chem. Soc.* **62**, 3275.
Fuller & Frosch 1939 *J. Amer. Chem. Soc.* **61**, 2575.
Levy & Corey 1941 *J. Amer. Chem. Soc.* **63**, 2095.
Lipson & Riley 1943 *Nature*, **151**, 250, 502.
Mack 1932 *J. Amer. Chem. Soc.* **54**, 2141.
Müller 1933 *Helv. Chim. Acta*, **16**, 155.
Schoon 1938 *Z. Phys. Chem.* **39B**, 385.
Senti & Harker 1940 *J. Amer. Chem. Soc.* **62**, 2008.
Warren & Bragg 1930 *Z. Kristallogr.* **76**, 201.

DESCRIPTION OF PLATES 1 AND 2

PLATE 1

FIGURES 6 to 11. X-ray diffraction photographs (copper $k\alpha$ radiation).

FIGURE 6. Drawn fibre.

FIGURE 7. Doubly-oriented strip, *c* axis parallel to X-ray beam. Plane of strip vertical.

FIGURE 8. Doubly-oriented strip, *c* axis vertical plane 75° to X-ray beam. Note that many of the reflexions of figure 1 are missing.

FIGURE 9. Doubly-oriented strip, *c* axis vertical plane parallel to X-ray beam. Taken on cylindrical film, axis horizontal.

FIGURE 10. Drawn fibre containing only α form.

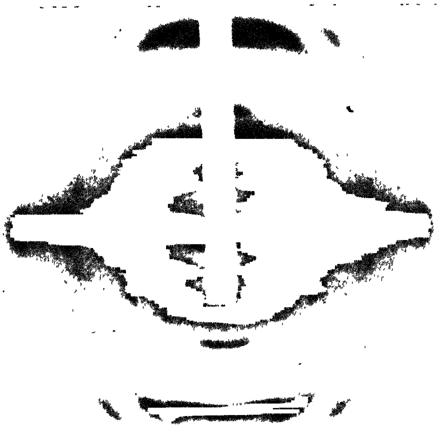
FIGURE 11. Ordinary drawn fibre. Monochromatic radiation. Flat film. Note absence of equatorial streaks.

PLATE 2

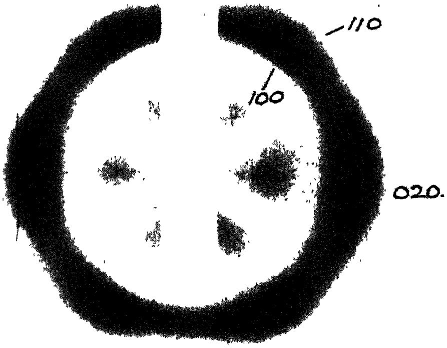
FIGURE 18. X-ray photograph of drawn fibre of 66 nylon, horizontal, oscillated 90° to 45° to X-ray beam.

FIGURE 19. X-ray photograph of quenched drawn fibre of 6.10 nylon.

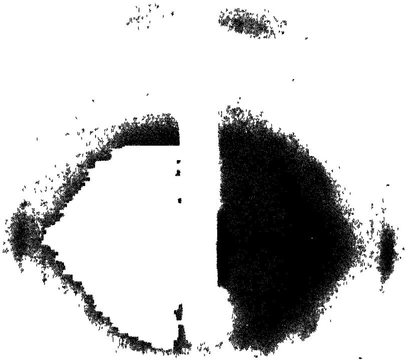
FIGURE 20. Conoscopic interference figures. *a*, 6.10 strip. *b* and *c*, different regions of a 66 strip.



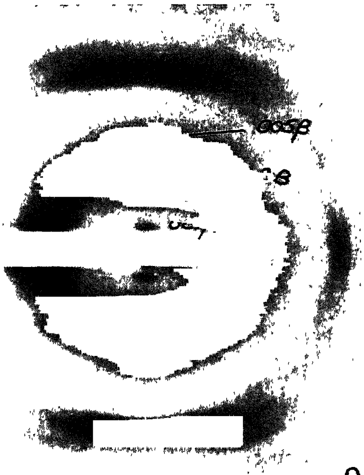
6



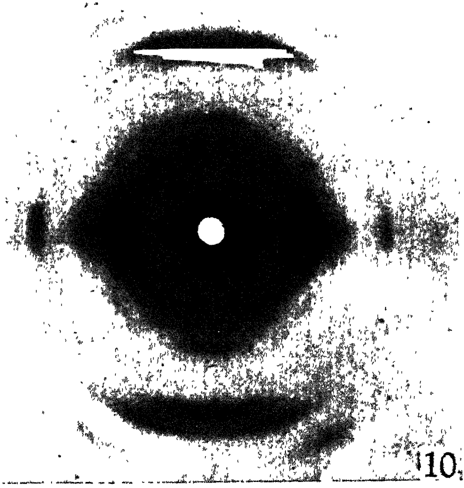
7



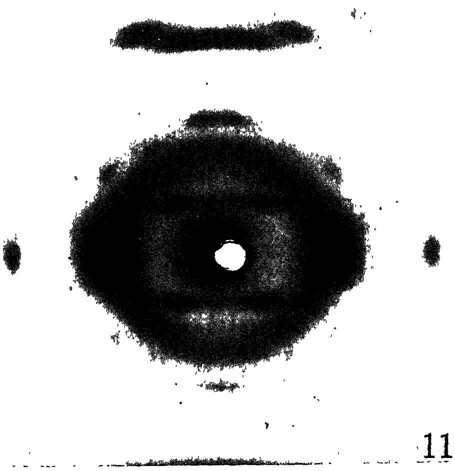
8



9

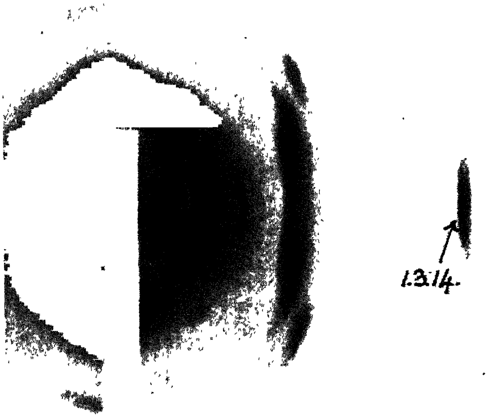


10

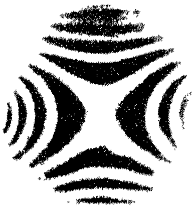


11

(Facing p. 68)

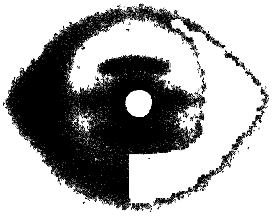


18

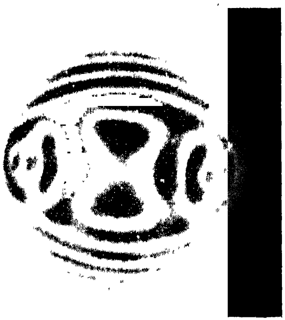


20*a*

20*b*



19



20*c*

Limitations of the Newtonian time scale in relation to non-equilibrium rheological states and a theory of quasi-properties

By G. W. S. Blair, D.Sc. and B. C. Veinoglou, Ph.D.
In collaboration with J. E. Caffyn, B.Sc.

National Institute for Research in Dairying, University of Reading

(*Communicated by E. N. da C. Andrade, F.R.S.—Received* 30 *May* 1946)

The behaviour of complex materials under stress is described in terms of entities which are not strictly 'physical properties'. These so-called 'quasi-properties' range from entities hardly distinguishable from dimensionally true physical properties to concepts which are much less clearly defined.

Quasi-properties measure an ordered process towards equilibrium rather than a state of equilibrium. The Newtonian definition for equality of time intervals which leads to the concepts velocity, acceleration, momentum and force having whole-number dimensional exponents, does not apply to 'quasi-equilibrium states'. In order to keep the Newtonian time scale, fractional differential equations are introduced.

The simplest fractional differential equation relating stress, strain and time integrates to a series equation whose first term is a simple power law (Nutting's equation) already known to describe the behaviour of many complex materials under constant stress. The physical meaning of the fractional differential is considered.

An apparatus is described for loading test-pieces of plastics and the like under tension or compression at constant stress to a preselected strain, and for following the subsequent stress dissipation; and the results of tests on thirty-eight materials are studied statistically.

Introducing a second term from the series equation (and, very rarely, a third term) greatly improves the fit for materials for which Nutting's equation is inadequate and explains hitherto unaccountable anomalies when the Nutting plot is otherwise satisfactory. Constants derived by the equation from constant-stress and constant-strain conditions are compared.

The form of the series equation suggests that the relative importance of the second term may sometimes disclose the presence of undissipated stresses in the materials.

The accuracy of tests on individual test-pieces is high, but, on account of frequent lack of homogeneity in the samples available, the use of the unmodified Nutting equation is often adequate even when the addition of a second term would significantly improve individual curves.

Some alternative treatments are discussed, but, both theoretically and practically, the fractional differential approach is preferred for most of the materials tested.

INTRODUCTION

It has been generally believed since the time of St Augustine (*Confessions*, Book xi) that, whereas space can be measured directly by the superposition of lengths, the measurement of time is indirect, being made either subjectively (as St Augustine thought, from memory engrams) or objectively, by definition, from some periodic physical process such as the rotation of the earth. St Augustine believed in an absolute time scale independent of the length of the day, and suggested that, even if the sun should be stayed in its course, time might still be measured by the rotation of the potter's wheel.

In Newtonian physics, times are implicitly defined as of equal length where either light or some body free from extraneous influences traverses equal (i.e. superposable) distances in them. It is assumed that lengths are themselves invariant. Recently, other phenomena have been found to give the same or a simply related time scale; e.g. Dingle (1944) equates the times taken by a specified body to radiate equal amounts of energy.

The rotations of the earth give only approximate agreement with the Newtonian scale and, as Dingle (1942) points out, astronomers prefer to say that the earth is slowing down rather than to question the accuracy of the Newtonian laws.

Bodies are said to move at constant velocity when they move through equal distances in equal times; but, since equal times are already defined as above, this is no more than the definition of an entity, velocity, being the first differential of length with respect to time. Acceleration is defined as the second differential, and the products with mass as momentum and force respectively.

In other words, equality of time units is so defined as to produce these simple whole-number differential concepts and, when a body does not, in fact, travel at constant rectilinear velocity, i.e. changes its momentum, force, which was originally a psycho-physical concept, is defined as the rate of change of the momentum. Einstein's relativity, while correcting the Newtonian laws for the special case of high velocities, does not affect the Newtonian time scale itself. Milne's relativity is not so easily dismissed, since no single time scale is regarded as universal and the measure of time takes priority over that of length. For the intervals of time with which we are concerned in the testing of materials, however, Milne's theories do not in any way affect our present line of argument.

From the Newtonian scale are derived all those physical properties which so conveniently give universal and generally integral indices of mass, length and time. But it has long been realized that the Newtonian time scale, in spite of its invaluable contribution to physics, remains an arbitrary scale. As Poincaré (1904) points out: 'La définition implicitement adoptée par les astronomes peut se résumer ainsi: le temps doit être défini de telle façon que les équations de la mécanique soient aussi simples que possible.' The Newtonian scale is the simplest for the study of most of the problems of physics.

For biological phenomena, however, du Noüy (1936) suggests an alternative time scale based on the rate of healing of aseptic wounds and points out that, since the units of this physiological time progressively increase in relation to the Newtonian time units, time will seem to pass more quickly as man ages. This raises an interesting parallel with Milne's theory, according to which the period of a pendulum shortens as we recede in time towards the finite 'beginning' of the atomic time scale.

This linking of physical, physiological and psychological time is especially interesting in view of recent experiments on the quantitative effects of temperature changes on the psychological time scale in man and in bees, ants, etc. (François 1928; Piéron 1945). Hoagland (1935) claims that the relationship between rate of counting and absolute temperature of a feverish patient is governed by a law of the Arrhenius type.

Bergson (1889) gives a wider definition of the equality of time intervals: 'Two intervals of time are equal when two identical bodies, in identical conditions at the beginning of each of these intervals and subject to the same actions and influences of every kind, have traversed the same space at the end of these intervals.' It will be noticed that this definition postulates the existence of bodies whose conditions are not permanently affected by the actions and influences during the specified time. According to Bergson's philosophy, living matter could never qualify.

Schrödinger (1944) describes living matter as '"doing something", moving...and so forth, and that for a much longer period than we would expect an inanimate piece of matter to "keep going" under similar circumstances'.

But there are many inanimate colloidal systems which behave in this way and hence do not qualify for Bergson's definition. In studying the behaviour of such systems one is at liberty either to use a non-Newtonian time scale, which is inconvenient since each system or group of systems has its own time scale; or, using the Newtonian definition, to make use of differentions of length with respect to time which are of fractional orders.

It will be seen that these alternatives are equivalent, since, as explained above, the definition of velocities and accelerations as first and second derivatives is effected by selecting the Newtonian definition of time equality. It is therefore proposed to keep to the Newtonian definition but to define two new generalized entities—motion, having dimensions $LT^{-(1+z)}$ and drive, with dimensions $MLT^{-(1+z)}$ in which z is a fraction.

In the experiments described below, these ideas were developed in relation to two processes: (1) the straining (tension or compression) of cylinders or strips of various colloidal materials—plastics and the like—under constant stress, and (2) the relaxation of stress of the same materials when subsequently held at constant strain. It is clear that the concepts of motion and drive need only be applied to conditions within the materials tested. Thus straining may be done at constant stress and not at constant drive per unit area, since the loads applied to the test-pieces are naturally to be treated in the ordinary Newtonian manner. Within the test-piece, however, the concept of stress must be regarded with caution.

In order to approach this rather difficult problem it is best first briefly to review the real significance of the rheological equations of the simple prototypes, the Hookean solid and the Newtonian liquid. (To avoid confusion, the latter will be referred to as the 'true liquid'.)

The simple Hookean system is generally a crystalline structure with regularly spaced inter-atomic potential troughs. The linear relationship between stress and strain within the small elastic range represents no more than an approximate cancellation of opposite curvatures from the power laws of attraction and repulsion. The most convenient entity by which to express the stress-strain relation is the rigidity modulus (n) ($ML^{-1}T^{-2}$), i.e. the simple ratio of (Newtonian) stress to strain.

Both the states of stress and of strain can be described by tensors made up of the three spatial components of the tractions and displacements (respectively) on each of the three faces of the unit cube. These ninefold tensors are symmetrical, for isotropic systems, so the only concern in this case is with six components. The appropriate vectors can be selected from these tensors to give the usual elastic constants, though in the work described in this paper volume changes can safely be ignored.

At the other end of the scale are the amorphous systems which tend to fluid behaviour. Here the reaction of the material may be regarded as uninfluenced by previous straining, so that the simple relationships between stresses and strain rates are independent of the magnitude of the strain and can thus be derived by a simple differentiation of the elasticity equations, for pure deformations.

The entity by which the stress-strain-time ($S : \sigma : t$) relation is best described is the ratio of shear stress to rate of shear (viscosity, η). Just as Hooke's law is no more than an approximation, so Grunberg & Nissan (1945) have shown that even for true fluids the viscosity is only constant under limited conditions.

Although there are some important materials which approximate to Hookean or to truly fluid behaviour, these are the exception rather than the rule. Between the crystalline and amorphous materials are many others whose structure is defined to a greater or lesser degree. Many of the so-called 'high-elastic' materials show approximately Hookean behaviour over large ranges of strain. The mechanism of this is quite different from that of ordinary elasticity, but the behaviour of the material may be described in terms of a simple ratio of stress to strain where this is found to be approximately constant. For really large strains, however, the strain axes do not lie at 45° with the direction of shear, so that although shear stress is proportional to shear strain, tensile stress is not proportional to tensile strain. Under these conditions, Mooney (1940) proposes a second modulus (for rubber) which is called the 'Coefficient of Asymmetry.'

Much work has been done on 'high elasticity', but in many otherwise valuable contributions to the subject it has been forgotten that the equations relating the tensile and shear moduli of elasticity are derived only by making certain assumptions such as that strains are small; and that these equations are not valid for large strains even when Poisson's ratio is $\frac{1}{2}$ (British Rheologists' Club 1946).

There still remain many materials—plastics, bitumens, pastes, unvulcanized rubbers, etc.—for which the inter-atomic potential troughs are very irregularly spaced and for which neither S/σ nor $S \div d\sigma/dt$ is constant. When these materials are strained, they are not, during the straining, in an equilibrium state in the usual sense of the term. By a system in equilibrium is generally meant one in which there are no unbalanced forces—force being here used in accordance with the Newtonian definition, although the term is also used less precisely to imply a balance of gains and losses, as when it is said that living matter maintains a state of equilibrium.

It is often necessary to study rheological behaviour in systems which are not in Newtonian equilibrium, and, as might be expected from what has already been said, it is found that the combinations of S, σ and t which are found to be constant do not prove to be physical properties in the sense that they do not have whole-number or even universal exponents of M, L and T; nor is the process by which they are measured reversible, in that the states through which the material passes are not equilibrium states, so that tests, though reproducible, must be repeated on fresh samples.

As typical of such entities take 'firmness', which, for many complex industrial materials is of great practical importance, can be reproducibly measured on replicate samples and in some cases by quite independent physical methods but which has dimensions which depend on the material or group of materials.

It has often been pointed out that such entities are not to be regarded as 'physical properties' in the strict sense of the term, and it has been suggested that, except for purely *ad hoc* industrial purposes, treatments involving them should be avoided. The alternatives, however, are far from satisfactory: they are, either to ignore such materials altogether or to presume that they are mechanical mixtures of simple Hookean and true-fluid prototypes or at least that they may be treated as such.

This latter assumption is occasionally justified, but although it is usually possible to reproduce an experimental rheological curve by fitting a number of exponential terms and always possible to invent a new prototype on conveniently exponential lines, neither process does much to further our knowledge of the behaviour of real materials. We propose, then, to accept the entities of the kind described above and to refer to them as 'quasi-properties'. They are a measure rather of a process than of a state, but it is a steady process, and, though not at equilibrium in the Newtonian sense, we feel justified in describing it as a state of 'quasi-equilibrium'. Thus the 'process' of work-hardening of a metal under pressure from a sphere in an indentation hardness test is measured by the Meyer exponent—there are many illustrations.

While it is true that the measurement of a 'property' must not be altered by the process of measurement (the axiom underlying Heissenberg's principle), there would seem to be no objection to measuring the process of change produced by the test provided that the methods are standardized and the results are independent of the observer.

It is clear that quasi-properties will be of different 'orders'. Some will hardly differ from true physical properties in that the exponents of MLT, although fractional, will be the same (within the limits of experimental error) for quite large groups of materials, and the quasi-properties themselves will be reproducible by widely different physical methods of measurements. At the other end of the scale will be quasi-properties of a low order which, although reproducible on replicate samples by some standard method, will neither be unique dimensionally for any appreciable range of materials nor equal quantitatively when derived from different types of test. Normally, one is faced with the choice of selecting a large number of true properties or a smaller number of quasi-properties, since it is nearly always possible to express rheological behaviour in terms of true properties provided enough constants are fitted.

It is hoped to show that the latter course is by no means always the best.

MATHEMATICAL TREATMENT OF QUASI-PROPERTIES. THE PRINCIPLE OF INTERMEDIACY

In many industries, the firmness of materials is assessed by handling by experts, and psycho-physical studies have been made of the thresholds (i.e. smallest perceptible differences) for various systems (Blair & Coppen 1939, 1940, 1942, 1943). It has also been found that experts handling materials tend to judge firmness not as if the materials were mixtures of elastic and viscous components but rather as though the quality of firmness lay between those of elastic modulus and of viscosity. This principle of ranking complex entities as lying at a certain distance between the prototypes, which we call the 'Principle of Intermediacy', common in psychology, is rare in rheology—one of the few existing examples in the latter science is the 'viscosity index' of an oil.

It may be objected that any such treatment will preclude any simple relation between the stress and strain tensors and will involve an infinite series of tensors. This is true, but, as has already been pointed out, for the very large strains which have to be considered, and, in view of non-equilibrium conditions, the usual tensor theory would be inapplicable in any case. Whether it will be possible to develop a new tensor theory, relating drive to strains, we do not know.

In the meantime, however, there are several possible applications of the Principle of Intermediacy which will help, bearing in mind that the materials given are in quasi-equilibrium and that there being no Newtonian bodies concerned within the material, the use of the second as the uniform time unit is sure to involve motions which are neither velocities nor accelerations. It is clear that in order to define a quasi-property it is necessary to specify (*a*) numbers indicating the dimensions of the quasi-property and (*b*) quantities measuring its intensity.

The first attempt was made by Blair & Coppen (1939) for constant stress conditions and was based on the following argument:

The entity which is found to be constant for true fluids is given by $\eta = S\sigma^{-1}t$; the entity which is found to be constant for elastic solids is given by $n = S\sigma^{-1}$. It was therefore proposed that the entity which might well be expected to be constant for complex materials would be

$$\psi = S\sigma^{-1}t^k, \tag{1}$$

where, normally, $1 > k > 0$.

This is a special case of a more general equation derived by intermediacy from the imaginary prototypes $S^0\sigma^{-1}t$ and $S^0\sigma^{-1}t^0$, for which we will redefine

$$\psi = S^\beta\sigma^{-1}t^k. \tag{2}$$

This equation had been empirically proposed by Nutting (1921*a*, *b*) many years earlier.

A more general treatment (Blair 1944) is to consider viscosity as defined by the equation $\eta = S \div d\sigma/dt$ and shear modulus as $n = S \div \sigma$.

A quasi-property χ_1 is then defined by the Principle of Intermediacy:

$$\chi_1 = S \div \left\lceil \frac{d^{\mu}\sigma}{dt^{\mu}}\right., \tag{3}$$

where $\lceil$ indicates a fractional differentiation. (We propose this convenient symbol since it reminds us of both fractions and Γ-functions.)

In our psycho-physical work we had already made use of the fractionally differentiated form of equation (2) for constant stress conditions:

$$\left\lceil \frac{\partial^{\nu}\sigma}{\partial t^{\nu}}\right. = \frac{\Gamma(k+1)}{\Gamma(k-\nu+1)} t^{k-\nu} \psi^{-1} S. \tag{4}$$

Our psycho-physical experiments did not at the time justify us in equating ν and k, but Caffyn & Blair (1945) later showed that *if* one is prepared to equate ν, k and μ (in equation (3)), the equations can be combined, since (4) becomes

$$\left\lceil \frac{\partial^{k}\sigma}{\partial t^{k}}\right. = \Gamma(k+1)\, \psi^{-1} S, \tag{4a}$$

from which it follows that

$$\frac{S}{\left\lceil \dfrac{\partial^{k}\sigma}{\partial t^{k}}\right.} = \chi_1. \tag{4b}$$

Hence

$$\psi = \Gamma(k+1)\, \chi_1. \tag{5}$$

Since $\Gamma(k+1)$ is a number, in this special case when the Nutting equation holds, ψ and χ_1 measure the intensity of the same quasi-property.

A second equation can also be obtained by fractional differentiation of stress with respect to strain

$$\chi_2 = \left\lceil \frac{\partial^{\rho} S}{\partial \sigma^{\rho}}\right. t, \tag{6}$$

which, taken together with equation (3), leads to the general equation

$$X = \left\lceil \frac{d^{\rho} S}{d\left(\left\lceil \dfrac{d^{\mu}\sigma}{dt^{\mu}}\right.\right)^{\rho}}\right.. \tag{7}$$

At present we are not able to make use of equation (7), but Bosworth (1946), quite independently of us, has recently proposed a somewhat similar quasi-property as a criterion of plasticity.

So far, the gamma function has not been much used in rheology. Its relation to power-law treatments is discussed by Gemant (1936, 1944) who has used half-differentials, but, although in his earlier paper he suggests that at least 'for secondary differential equations [the fractional differential] may really occur in a physically realizable case', in his later paper he expresses the view that it 'only occurs as a useful mathematical symbol, whereas the underlying elementary process, whatever it may be, will probably contain differential quotients of an integral order'. His view of the matter is clearly very different from ours. Our own view of the physical meaning of the fractional differential is somewhat as follows:

If cylinders of Californian bitumen (which is very nearly a true fluid) are compressed under constant stress, the amount of deformation produced will depend equally on the stress and on the time of its application. In the case of rubber, which may be regarded as nearly Hookean, the strain will depend on the stress only and not at all on the time. In the case of a certain plastic, strain is found to depend on stress and a fifth as much on the time, i.e. the Nutting equation holds and $k = 0{\cdot}2$. When this simple equation does not hold for constant-stress experiments, the next simplest hypothesis is that, just as there is a constant property for the bitumen (viscosity) which is given by the ratio of stress to the first differential of strain with respect to time, and a constant ratio (shear modulus) for the rubber given by the stress divided by the zero-differential of strain with respect to time, so the plastic has a quasi-property which is constant and given by the ratio of stress to the 0·2 differential of strain with respect to time. It must be remembered, of course, that if this plastic were the only physical system to be considered, it would be simpler to use a different rule for equating time intervals chosen so that the constant entity would have dimensions with whole-number exponents.

This view has strong experimental support (Blair & Coppen 1943), since it is found that in a statistical study of the compression of a large number of cylinders of a plastic in direct handling comparison with cylinders of rubber or bitumen, k, calculated from the relative importance of squeezing time, agrees (within experimental error limits) with k derived from simple compressions on a machine at constant stress. We regard this experimental fact as strong evidence for the validity of our line of approach.

It is known that some materials show a simple power-law relation between stress and strain not involving time (Bach's law), though as Reiner (1933) points out this does not involve us in dimensional difficulties. There are, on the other hand, many systems for which β (in equation (2)) is almost unity, though we have shown this to be, in some cases, a chance cancellation of the opposite effects of plasticity and a type of dilatancy (Blair & Veinoglou 1944*a*).

All these phenomena may not be found to be explainable in terms of the non-Newtonian time scale, but certainly many of the anomalies hitherto unexplained would seem to be clarified by it.

Although we have shown (and indeed so have many other authors) that a very large number of materials obey the simple Nutting law within the limits of experimental error, it is certainly worth while to consider what happens when the situation is not so simple.

Equation (4) can be integrated to give an equation of the type

$$\sigma = S^{\beta}(At^{k'} + Bt^{k'-1} + Ct^{k'-2} \ldots), \tag{8}$$

and it is clear that the Nutting equation is a special case of this when $A \gg B, C, \ldots$, etc., and that the addition of subsequent terms would not only account for differences between k and μ but might also express the quasi-properties of those of our materials whose agreement with the simple Nutting equation is unsatisfactory. In fact we will show in the experimental part that it is rarely necessary to go beyond the second term.

In the light of this introduction we would then summarize the purposes of our experiments as follows:

We believe that in the case of many complex colloidal materials, such processes as simple straining (under constant stress) and stress dissipation (under constant strain) are not equilibrium processes in the Newtonian sense, and that the time system which would most simply explain their behaviour is not Newtonian time. If Newtonian time is used, the entities which most simply describe their behaviour, which we call 'quasi-properties', will not have whole-number dimensions in *MLT*, and their fractional dimensions will be more or less dependent on the nature of the material. We would ask ourselves: (1) With what accuracy can these 'quasi-properties' be measured in straining and relaxation and how will their values obtained by these two very different methods agree with one another? (2) Simple tensor theory is not applicable to these conditions, nor has the necessarily more complex non-linear tensor theory been developed; but we have proposed certain fractional differential equations from which the 'quasi-properties' associated with 'firmness' can be defined. How well do these equations fit the experimental data, and how does the fit and the simplicity of treatment compare with the classical methods of deriving true 'physical properties'?

EXPERIMENTAL

Data of a kind suitable to test the theory outlined above are best obtained by holding either stress or strain constant and measuring the rate of change of the remaining variable with time. There are various ways of extending strips or rods in such a way that the load diminishes proportionally to the average change in cross-section of the test-piece (volume changes being ignored), but if a single instrument is to be used both for extensions and compressions, there is everything to be said for the bent lever type of compensator. The limitations and uses of this type of constant-stress device has been worked out fully by Caffyn (1944). Blair & Veinoglou (1944*b*) have already shown that many complex colloidal systems tested on machines of this type follow the Nutting equation with reasonable accuracy. In preliminary experiments a few materials were tested after rapid straining in relaxation at constant strain as well as by comparatively slow straining at constant stress.

Equation (2) can be partially differentiated for these two processes to give $\frac{\partial S}{\partial t} = -\frac{k}{\beta}\frac{S}{t}$ for relaxation and $\frac{\partial \sigma}{\partial t} = k\frac{\sigma}{t}$ for straining. Thus the slope of the relaxation curve plotting $\log S/\log t$ is k/β, and of the straining curve, plotting $\log \sigma/\log t$ is k. The constant β can be calculated from the position of a series of curves obtained at different strains and stresses respectively.

For the few materials tested in this way it was found that the Nutting constants k, β and ψ were not only remarkably constant in each experiment but also their values obtained in relaxation and in straining were the same within the limits of experimental error.

But to make a thorough study of these materials an apparatus was made by which a strip or rod (*B*) could first be extended at constant stress, the load being compensated for average change in cross-section by means of a bent lever (*H*) passing from the horizontal downwards towards the vertical; or to compress cylinders (*A*) with correctly increasing load effected by a lever moving from the vertical towards the horizontal (*G*). The movement is checked at a preselected strain by means of collars set the correct distance apart (at *J*). The subsequent stress dissipation is measured by a strong spring (*C*) resting on knife-edges, serving as a base for the compressed cylinder or a ceiling to which the extended strip is attached. The movements of the spring are negligible by comparison with the deformations produced in the test-piece, and the load on the spring is measured by an optical lever (*D*) which magnifies some thousandfold.

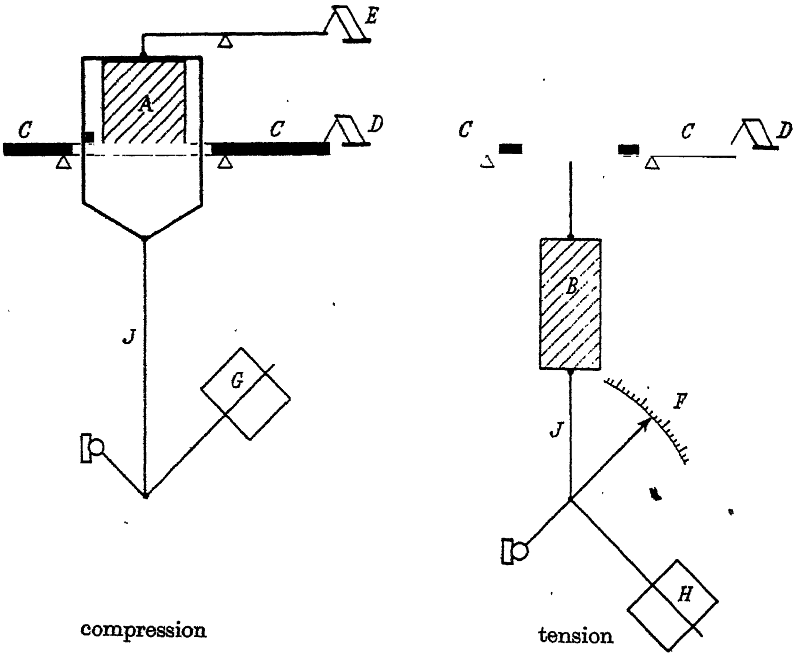


FIGURE 1. Compensated straining and relaxation apparatus.

The deformations of the stretched strip are measured by a pointer on a scale (*F*), but those of the compressed cylinder are too small for this and are suitably magnified by a second optical lever (*E*) operating from the stirrup which loads the cylinder. The whole instrument, which is shown diagrammatically in figure 1, is operated in a constant-temperature room, some experiments being done at 15° C and some at 18° C.

With this apparatus, test-pieces could be strained under high or low stress and to large or small preselected strains and the subsequent stress dissipations followed.

Since 'quasi-properties' refer to processes rather than to states, the question arises whether relaxation should be regarded as a separate process from straining, a new time zero being taken; or whether a single time scale should be used throughout the double experiment. For very rapid straining the difference between these two methods is negligible and the best relaxation data were obtained in this way, being compared with data from separate and slower straining experiments. For relaxation following very slow straining it is best to use a single time scale. If a new time zero is taken there will be an appreciable period before the new state of quasi-equilibrium is established and, on account of frictional effects, the instrument is not suited to very slow working. Finally, all our data were calculated by the two-process method, a small number of data from slow straining experiments being discarded.

As already pointed out, altogether apart from our treatment, the classical elasticity equations relating tensile to shear conditions are not valid for the large strains in which we are interested, nor do our materials generally remain isotropic on straining.

It would really have been better if, from tensile stresses and strains, 'tensile' constants had been derived even though shear modulus has more fundamental significance than Young's modulus; but, historically speaking, the constant ψ in the Nutting equation has always been defined as analogous to viscosity and shear modulus rather than to coefficient of viscous traction and Young's modulus. For this reason, ψ is calculated from *a* shear stress taken as $\frac{1}{3}$ of the tensile (or compressive stress). This is purely a matter of convention and has no theoretical meaning. The factor 3 could be replaced by $2(1+\Pi)$, but Π (Poisson's ratio) differs little from $\frac{1}{2}$ for these materials.

There are various ways of defining large strains. Like Bridgman (1945) we prefer the natural strain, $\log_e(l/l_0)$ and have used this formula throughout. It has the great advantage of being additive.

When possible, straining curves were obtained at some six constant stresses and relaxation curves starting from about four different initial strains for each of these stresses. A complete test on a material thus comprised about twenty-four curves, though in many cases such a complete experiment was not possible owing to shortage of materials.

Curves from part of a typical experiment with a thiokol are given in figure 2. On the left-hand side of the figure, $\log\sigma/\log t$ curves are given for seven different constant stresses; and above is the derived curve plotting $\log S$ against $\log\sigma_{10}$ where σ_{10} is the strain after 10 sec. The slope of this curve, by the method of least squares, gives β.

The part of the relaxation experiment shown on the right of the figure corresponds to six different constant strains, all the samples being submitted to the same initial stress ($\log S_0 = 6{\cdot}07$). There were other curves at other stresses, but to show more in a single figure would be confusing. Above are plotted the six corresponding points

$\log \sigma/\log S_{10}$, where S_{10} is the stress after 10 sec. dissipation. The slope of the lower curves gives k/β and that of the upper curve, though of course with many more points, gives β in relaxation. (We are not the first to apply power-laws to relaxation of stress. Prache (1930) gets excellent curves for unvulcanized rubber.)

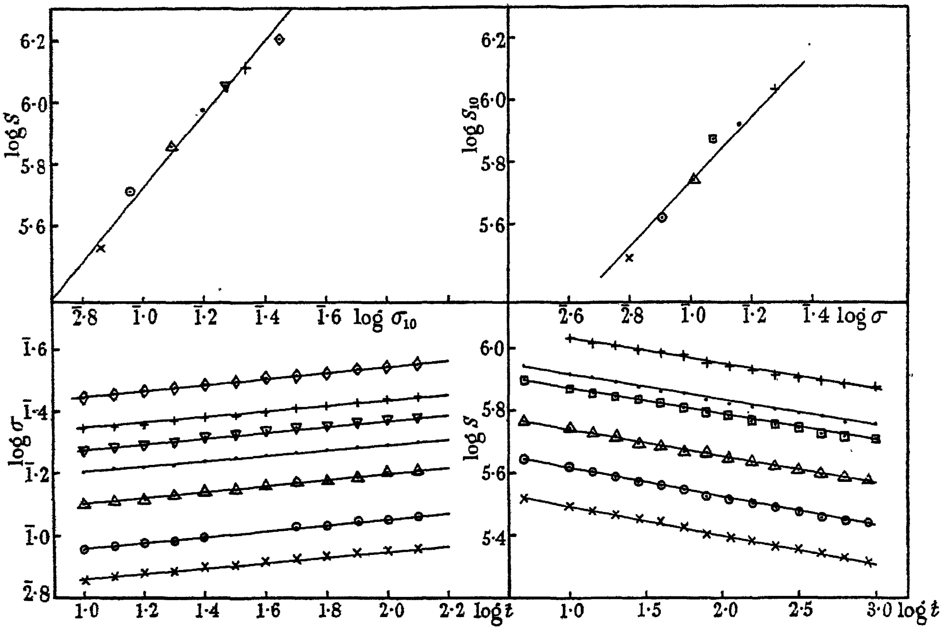


FIGURE 2. Compression and relaxation curves for a typical thiokol. S is in dynes/cm.², t is in sec.

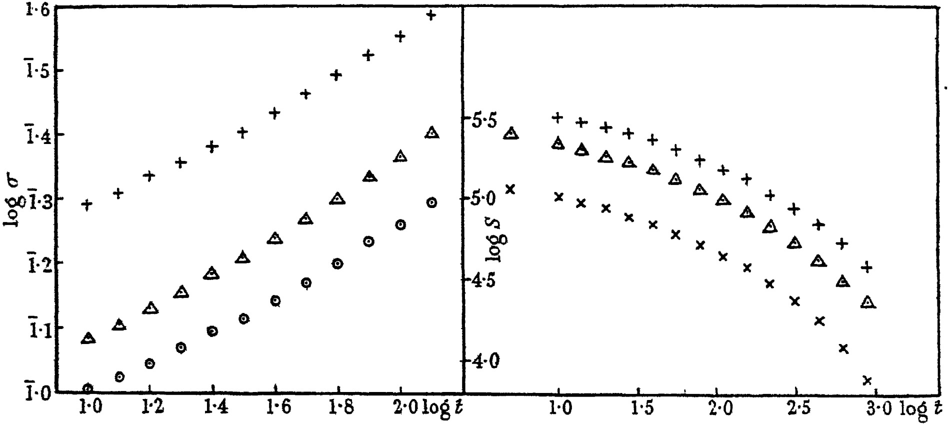


FIGURE 3. Nutting plots for compression and relaxation of highly anomalous materials.

compression	relaxation
⊙ $S = 17{\cdot}8 \times 10^4$	× $\sigma = 0{\cdot}0624$
△ $S = 21{\cdot}9 \times 10^4$	▽ $\sigma = 0{\cdot}146$
+ $S = 30{\cdot}2 \times 10^4$	+ $\sigma = 0{\cdot}263$

These curves are shown as being reasonably typical. The Nutting treatment (without addition of a second term) in this case is not particularly satisfactory in relaxation.

Unless the discrepancies are exceptionally large it is not possible to show clearly the effects of fitting the second term graphically or even in tabular form. A rather extreme case of a butyl rubber is given in table 1, the Nutting plots of which are shown in figure 3. To economize space, only three stresses are shown.

TABLE 1. COMPARISON OF EQUATIONS (2) AND (8) WITH EXPERIMENTAL DATA FOR COMPRESSION AND RELAXATION FOR A HIGHLY ANOMALOUS BUTYL RUBBER

compression

$S \times 10^4$ ...	17·8			21·9			30·2		
t	σ_8	σ_2	$\sigma_{\text{exp.}}$	σ_8	σ_2	$\sigma_{\text{exp.}}$	σ_8	σ_2	$\sigma_{\text{exp.}}$
10	0·1016	0·0984	0·1012	0·1217	0·1210	0·1216	0·1963	0·1906	0·1959
13	0·1057	0·1046	0·1059	0·1276	0·1264	0·1274	0·2047	0·2029	0·2042
16	0·1107	0·1111	0·1109	0·1345	0·1351	0·1349	0·2149	0·2159	0·2173
20	0·1165	0·1180	0·1175	0·1426	0·1444	0·1429	0·2269	0·2298	0·2275
25	0·1233	0·1254	0·1245	0·1518	0·1544	0·1531	0·2404	0·2445	0·2410
32	0·1308	0·1332	0·1303	0·1621	0·1650	0·1622	0·2557	0·2603	0·2541
40	0·1401	0·1417	0·1387	0·1737	0·1764	0·1738	0·2726	0·2770	0·2723
50	0·1486	0·1505	0·1497	0·1864	0·1886	0·1862	0·2912	0·2948	0·2917
63	0·1589	0·1599	0·1585	0·2003	0·2016	0·1995	0·3116	0·3137	0·3126
79	0·1701	0·1699	0·1718	0·2156	0·2155	0·2163	0·3339	0·3339	0·3350
100	0·1823	0·1805	0·1820	0·2323	0·2303	0·2328	0·3582	0·3354	0·3589
126	0·1955	0·1918	0·1972	0·2504	0·2462	0·2523	0·3846	0·3782	0·3864

relaxation

	$\sigma = 0{\cdot}0624$			$\sigma = 0{\cdot}146$			$\sigma = 0{\cdot}263$		
t	S_8	S_2	$S_{\text{exp.}}$	S_8	S_2	$S_{\text{exp.}}$	S_8	S_2	$S_{\text{exp.}}$
10	9·926	12·050	10·420	21·410	25·500	22·390	31·180	37·350	32·430
14	9·800	10·340	9·550	20·980	22·020	20·510	31·000	32·580	30·480
20	9·147	8·861	8·810	19·560	19·010	18·710	29·280	28·420	28·050
28	8·212	7·600	7·762	17·600	16·420	16·980	26·650	24·790	25·640
40	7·161	6·518	7·079	15·440	14·180	15·210	23·620	21·620	23·390
56	6·130	5·590	6·166	13·300	12·250	13·300	20·560	18·870	20·700
79	5·171	4·793	5·248	11·320	10·570	11·610	17·660	16·450	18·030
112	4·319	4·111	4·519	9·543	9·128	10·000	15·030	14·350	15·700
159	3·583	3·526	3·890	7·997	7·883	8·375	12·720	12·520	13·270
224	2·958	3·024	3·048	6·670	6·808	6·934	10·700	10·920	10·940
316	2·431	2·592	2·421	5·547	5·878	5·495	8·976	9·530	9·036
447	1·995	2·223	1·820	4·603	5·076	4·266	7·515	8·316	6·982

σ_8 is strain and S_8 is stress ($\times 10^4$) calculated from two terms of equation (8).
σ_2 and S_2 are stress ($\times 10^4$) calculated from equation (2) (Nutting's law).
$\sigma_{\text{exp.}}$ and $S_{\text{exp.}}$ are experimental values.
t is time in sec. Initial stress in relaxation (all three curves) = $39{\cdot}3 \times 10^4$ dynes/cm.^{2}.

The relaxation data illustrate one of those rare cases where even the addition of a second term, although effecting a great improvement, still leaves something to be desired. The addition of a third term from equation (8) improves the fit still further, but this improvement would not normally justify the large amount of labour involved in the calculation.

In introducing a second or even a third term of equation (8), it is important to bear in mind their significance over different parts of the time scale. For very short times when A and B have the same sign, the two-term plot goes through a minimum in straining or a maximum in relaxation since the second term with a negative exponent preponderates. This has no more physical significance than the minimum in the well-known Buckingham-Reiner equation if it is applied to stresses below the yield-value (see Blair 1938). The condition of quasi-equilibrium which the treatment postulates has not been reached at such small values of t.

But it is curvature in the Nutting curves for comparatively short times that is most effectively corrected by introducing the second term.

A good Nutting curve which deteriorates only after long times is not easy to deal with by means of equation (8). Downward curvature of this type is rare, and upward curvature suggests an examination of the test-piece for signs of rupture or, in the case of extensions, slipping at the grips.

In relaxation, where deviations from simple Nutting behaviour are on the whole more serious, curvature is almost invariably found in the short-time part of the curves.

There are a number of cases where the $\log \sigma/\log t$ plot for a series of constant stress experiments gives curves which superficially appear to be linear, but it is found that the slope, k, increases or decreases progressively with increasing stress. It is thought likely that this is due to the fact that the Nutting equation is only the first term of the series equation (equation (8)), and that even though the second term correction might be small, the value of k, determined over the experimental range, might differ significantly from the true slope, i.e. the slope at large strains. If this is the explanation, then any slight upward curvature of the $\log \sigma/\log t$ curve should lead to an apparent increase in k with increasing stress, whereas downward curvature would lead to the reverse. Curvature is easily checked by comparing the magnitude of k from the Nutting equation and k' from the two-term fit, and it was found in the great majority of cases that any tendency for k to vary with stress was in the sense predicted on this hypothesis. Evidence is, however, not adequate to state categorically that all apparent variations of k with stress are due to this cause.

As well as the problem of fixing a time zero, there is also the question of the strain zero. For simple elastic systems this is easy, but for complex systems zero strain means no more than the shape of the test-piece at the start of the experiment, assuming that the material is then in the normal rest condition with which the test is required to deal. As Love (1906) puts it: 'The unstrained state...is, as it were, an arbitrary zero of reckoning, and the choice of it is in our power.'

The presence of undissipated internal stresses, which would lead to the fixing of a zero of strain different from that selected if the material were devoid of internal stresses, would produce a curvature in the $\log \sigma/\log t$ curves assuming a normal Nutting behaviour for the unstressed system. Furthermore, the source of 'error' would lead to a modification of the Nutting equation precisely the same as equation (8).

This suggests that the presence of a significant second term in equation (8) may, at least in some cases, be indicative of undissipated internal stresses, a point which, if verified, would be of considerable practical importance.

Unfortunately, we have neither time nor suitable materials to test this hypothesis adequately, though preliminary experiments are sufficiently encouraging to warrant further investigation.

ALTERNATIVE TREATMENTS

As pointed out in the introduction, it is possible, by postulating enough exponential terms, to obtain an adequate fit to almost any S/t curve. In order to get accuracy comparable to that derived from the two-term form of equation (8), either two or three exponential terms are generally required. The equations, which were calculated by Whitehead's (1935) method, are of the form $S = pe^{-t/\tau_1} + qe^{-t/\tau_2} + r$ which means that a two-term exponential involves five parameters ('properties') and a three-term involves seven. The work involved depends on the method used. With Whitehead's graphical method, the calculation takes less time than that for the two-term equation (8), but it must be remembered that the latter calculation gives a statistical fit whereas the former does not give the errors in the fitted parameters.

As a second alternative, Alexandrov & Lazurkin's (1940) method was tried, though in the majority of cases the assumptions that there is an easily recognizable 'immediate' elastic reaction σ_0 and that, under constant stress, the strain approaches asymptotically to an equilibrium (maximum) value (σ_∞), are so wide of the mark for most of the materials tested by us that it is rarely possible to get any fair measure of σ_0 or σ_∞. For elastic bitumen, for which these constants could be determined with some confidence, the fit of the Alexandrov-Lazurkin equation was definitely bad.

Apart from the earlier work already published, we had some 173 straining and 455 relaxation curves suitable for a complete statistical treatment using equation (8). These data cover some thirty-eight 'materials', counting characteristically different members of the same family of polymers as distinct 'materials'.

STATISTICAL TREATMENT

Under constant stress conditions, equation (2) gives linear $\log\sigma/\log t$ curves, and under constant strain conditions linear $\log S/\log t$ curves. The constants of these equations have been determined by fitting a straight line to the data by the method of least squares. It was assumed that the $\log t$ values were known accurately in both cases and the error was in $\log\sigma$ and $\log S$ respectively.

The constants of equation (8), (for tension) using the first two terms, were determined by transforming it to

$$\log\sigma = \beta\log S + \log A + k'\log t + \log\left[1 + B/At\right]$$

or

$$\log\sigma = \beta\log S + \log A + k'\log t + BM/At$$

to the first approximation, provided $BM/A < 1$, $t > 1$ and $M = \log_e 10$. This equation can be written

$$y = a + bx + c\times 10^{-x},$$

where $y = \log\sigma$, $x = \log t$, $a = \log S + \log A$, $b = k'$, $c = BM/A$.

The values of a, b and c were determined by the method of least squares, assuming x to be known exactly.

A similar transformation can be made for constant-strain conditions (relaxation) and the constants determined in the same way.

The summarized results of this analysis are given in table 2.

Discussion of results

In the case of the simple Nutting equation, the constant k is directly derived from the rate of straining of a single test-piece under constant stress, and the rate of dissipation of the same test-piece gives k/β. The calculation of β, on the other hand, involves data obtained from a number of test-pieces, and there is no doubt that the very much higher standard error in β is due to the fact that lack of homogeneity of the samples from which the test-pieces were taken is very much the limiting factor in defining the accuracy of comparison between different materials. For this reason it can hardly be over-emphasized that when it is stated that the introduction of a second term significantly improves the fit for a given curve, that does not mean that the unmodified Nutting equation would be inadequate if the purpose of the tests is to compare one material with another. The adequacy of any equation for this purpose depends on the homogeneity of particular materials. If two strips from the same sheet of plastic differ appreciably, as is often the case, there will be little point in improving on the Nutting plot in order to get effective constants for the material unless the second term is very much larger than that which would necessitate its use if only individual curves were considered.

Having given this warning it is now possible to proceed to discuss the proportion of cases in which *individual curves* can be improved by the use of a second term.

In the straining experiments (tension and compression) 66 curves out of 173 (38 %) show significant improvement on fitting a second term. In terms of materials

47 % of materials are significantly improved in less than 25 % cases,
24 % of materials are significantly improved in 25 to 75 % cases,
29 % of materials are significantly improved in more than 75 % cases.

In all cases where there was no significant 'improvement' it is implied that the Nutting fit was adequate within the limits of experimental error. There were no

TABLE 2. MEAN NUTTING CONSTANTS OF MATERIALS AND SIGNIFICANCE OF ADDITION OF A SECOND TERM

	straining								relaxation						
Material	number of curves showing significant curvature	number of curves showing insignificant curvature	percentage showing significant curvature	k	standard error of $k \times 10^2$	β	standard error of $\beta \times 10^2$	$\frac{\% \text{ error of } k}{\% \text{ error of } \beta}$	number of curves showing significant curvature	number of curves showing insignificant curvature	percentage showing significant curvature	k/β	standard error of $k/\beta \times 10^2$	β	k
polyvinyl chloride (1)	8	2	80	0·116	0·39	0·56	8	0·23	17	8	68	0·122	0·31	0·89	0·108
polyvinyl chloride (2)	0	1	0	0·108	0·33	0·96	8	0·37	7	1	87	0·066	0·73	0·96	0·063
polyvinyl chloride (3)	1	0	100	0·119	0·44	0·80	4	0·74	12	3	80	0·111	0·35	0·87	0·097
polyvinyl chloride (4)	2	1	66	0·037	0·15	0·71	3	0·96	6	7	46	0·064	0·19	0·72	0·046
polyvinyl chloride (5)	2	1	66	0·111	0·27	0·74	8	0·22	11	8	52	0·119	0·28	0·84	0·100
polyvinyl chloride (6)	1	1	50	0·098	0·28	0·85	3	0·81	11	4	73	0·108	0·30	0·78	0·084
polyvinyl chloride (7)	2	1	66	0·380	1·10	1·80	—	—	1	3	25	0·241	0·72	1·66	0·400
polyvinyl chloride (8)	3	7	30	0·075	0·22	0·78	3	0·76	12	10	54	0·095	0·26	0·93	0·890
polyvinyl chloride (9)	2	7	22	0·096	0·27	0·93	3	0·87	8	11	42	0·099	0·20	1·05	0·104
plastic (unspecified 1)	—	—	—	—	—	—	—	—	3	3	50	0·089	0·14	0·96	0·086
plastic (unspecified 2)	2	0	100	0·289	0·86	0·13	—	—	4	2	66	0·298	1·30	—	—
plastic (unspecified 3)	2	0	100	0·278	0·70	0·90	—	—	2	3	40	0·277	0·53	—	—
plastic (unspecified 4)	3	0	100	0·566	1·40	1·00	—	—	8	1	89	0·372	3·30	—	—
plastic (unspecified 5)	1	2	33	0·737	1·11	0·90	—	—	3	0	100	1·070	9·60	0·45	0·482
cellulose acetate + phthalate	0	1	0	0·436	0·49	2·40	—	—	7	0	100	0·277	0·90	3·70	1·023
rubber—unvulcanized (1)	3	0	100	0·118	0·46	0·52	13	0·16	5	1	83	0·188	0·50	0·94	0·177
rubber—unvulcanized (2)	5	0	100	0·223	0·53	0·84	1	1·99	4	4	50	0·199	0·43	1·47	0·293
rubber—sulphur-free (3)	0	8	0	0·022	0·13	1·07	2	3·15	3	19	14	0·038	0·09	1·19	0·045
elastic bitumen (1)	0	2	0	0·164	0·51	1·16	16	0·23	—	—	—	—	—	—	—
elastic bitumen (2)	0	2	0	0·121	0·30	0·82	11	0·18	4	10	29	0·126	0·30	0·89	0·112
bitumen (1)	7	1	88	0·450	0·88	0·80	—	—	14	0	100	0·390	2·40	1·00	0·390
bitumen (2)	3	0	100	0·429	1·20	0·90	—	—	10	0	100	0·257	1·20	—	—
chewing-gum	4	0	100	0·676	1·05	0·76	11	0·11	—	—	—	—	—	—	—
chewing-gum base	1	2	33	0·373	0·89	1·17	24	0·12	—	—	—	—	—	—	—
rye-flour toffee	0	2	0	0·962	0·67	2·07	50	0·03	—	—	—	—	—	—	—
milk toffee (gelatine base)	2	0	100	0·737	1·90	1·00	—	—	—	—	—	—	—	—	—
'Hycar'	0	2	0	0·082	0·35	1·51	7	0·92	11	3	78	0·054	0·15	1·79	0·097
'butyl' (1)	0	4	0	0·037	0·19	1·37	12	0·59	0	8	0	0·036	0·21	1·70	0·062
'butyl' (2)	6	0	100	0·275	0·65	2·07	50	0·10	21	0	100	0·423	2·50	—	—
neoprene A	0	4	0	0·034	0·52	1·13	3	5·86	2	6	25	0·045	0·09	1·22	0·055
neoprene B	0	5	0	0·021	0·19	1·03	6	1·55	2	4	33	0·026	0·09	0·95	0·025
neoprene E	1	3	25	0·036	0·17	1·00	8	0·59	4	3	57	0·041	0·10	1·06	0·044
neoprene F	2	4	33	0·029	0·19	1·16	4	1·90	3	6	33	0·035	0·11	1·20	0·042
thiokol F.A.	0	16	0	0·095	0·24	0·88	2	1·11	7	20	26	0·088	0·15	0·75	0·067
thiokol A.Z.	3	4	43	0·121	0·26	0·91	2	0·98	22	0	100	0·144	0·61	1·00	0·144
polyisobutylene	0	6	0	0·134	0·28	1·56	5	0·65	8	12	32	0·088	0·14	2·78	0·246
G.R.S. (synthetic rubber)	0	8	0	0·085	0·19	1·29	2	1·45	7	20	26	0·075	0·14	1·43	0·105
Cheddar cheese	0	4	0	0·190	0·30	1·34	—	—	15	5	75	0·176	0·41	2·72	0·479
skim-milk cheese	0	6	0	0·190	0·21	0·91	—	—	16	0	100	0·216	0·66	—	—

cases of bad Nutting fit where the introduction of a second term did not effect an improvement and only a very few (and then only in relaxation, e.g. see table 1) where the fitting of one further term did not give a statistically satisfactory fit.

In the relaxation experiments, the simple Nutting equation was less adequate than in straining. Thus, of 455 relaxation curves analysed, 270 (59 %) were significantly improved by the addition of a second term.

Only 9 % of materials tested required a second term in less than 25 % cases,
54 % of materials tested required a second term in 25 to 75 % cases,
and 37 % of materials tested required a second term in more than 75 % cases.

It will be recalled that in the preliminary relaxation experiments in which the initial straining was very rapid, for the few materials tested, values of ψ, k and β were strikingly similar when obtained by straining at constant stress and by relaxation. With the slow controlled pre-straining of the later experiments, there were, in general, greater divergencies. Confining our comparisons only to materials for which the unmodified Nutting equation holds adequately, some results are given in table 3.

TABLE 3. COMPARISON OF CONSTANTS (MEANS) OBTAINED IN STRAINING AND RELAXATION

material	straining log ψ	relaxation log ψ	straining k	relaxation k	straining β	relaxation β
sulphur-free rubber	7·76	8·56	0·022	0·046	1·07	1·19
elastic bitumen	5·82	6·46	0·121	0·112	0·82	0·89
butyl rubber	9·44	11·65	0·037	0·062	1·37	1·70
neoprene A.	8·24	8·88	0·034	0·056	1·13	1·22
thiokol F.A.	6·05	5·44	0·095	0·067	0·88	0·75
G.R.S. (synthetic rubber)	9·49	10·40	0·085	0·105	1·29	1·43

CONCLUSIONS

We conclude that for the practical comparison of the rheological behaviour of many materials of the kind which we have studied, the Nutting treatment is not only adequate and relatively simple, but expresses those quasi-properties of materials which are often important in industry and are certainly used extensively for assessing their quality.

The limiting factor at present is lack of homogeneity of samples, and, in some cases, latent internal stresses. Where these difficulties can be overcome, it is worth while in a fair proportion of cases to extend the Nutting treatment by introducing the second term in the series equation of which it forms a part. Even then, the properties of the material are described by only four constants, i.e. less than the five constants required for even a two-term exponential treatment.

On the theoretical side much remains to be done, but we are convinced that a sound start has been made in approaching the problem in terms of the fundamental definition of the equality of time intervals.

The use of constant-strain relaxation as a method of studying the rheology of plastics is open to various objections. It is likely that the method will always be secondary to direct straining methods, but we feel that it has its uses, especially when it is desirable to measure properties or quasi-properties by two quite independent methods.

We would express our best thanks to the British Electrical and Allied Industries Research Association for permission to publish this work and to many friends, too numerous to mention by name, who have supplied us with materials and helpful advice. But we would especially thank Dr S. Whitehead, Dr P. White and Professor H. Dingle for their help and encouragement, and Miss M. Baron, Mr W. Lethersich and Miss J. Hazeldean for carrying out many statistical calculations.

REFERENCES

Alexandrov, A. P. & Lazurkin, J. S. 1940 *Acta Physicochim. U.R.S.S.* **12**, 647.

Bergson, H. 1889 *Essai sur les données immédiate de la conscience*, p. 115. Transl. Pogson (*Time and free will*), London, 1910. Swan Sonnenschein and Co.

Blair, G. W. S. 1938 *Introduction to industrial rheology*. London: J. and A. Churchill.

Blair, G. W. S. 1944 *J. Sci. Instrum.* **21**, 80.

Blair, G. W. S. & Coppen, F. M. V. 1939 *Proc. Roy. Soc.* B, **128**, 109.

Blair, G. W. S. & Coppen, F. M. V. 1940 *Brit. J. Psychol.* **31**, 61.

Blair, G. W. S. & Coppen, F. M. V. 1942 *Amer. J. Psychol.* **55**, 215.

Blair, G. W. S. & Coppen, F. M. V. 1943 *Amer. J. Psychol.* **56**, 234.

Blair, G. W. S. & Veinoglou, B. C. 1944*a* *J. Sci. Instrum.* **21**, 149.

Blair, G. W. S. & Veinoglou, B. C. 1944*b* *Nature*, **153**, 165.

Bosworth, R. C. L. 1946 *Nature*, **157**, 447.

Bridgman, P. W. 1945 *Rev. Mod. Phys.* **17**, 3.

British Rheologists' Club (Committee) 1946 *Essays in Rheology*. London: Sir Isaac Pitman and Sons.

Caffyn, J. E. 1944 *J. Sci. Instrum.* **21**, 213.

Caffyn, J. E. & Blair, G. W. S. 1945 *Nature*, **155**, 171.

Dingle, H. 1942 *Phil. Mag.* **33**, 321.

Dingle, H. 1944 *Phil. Mag.* **35**, 499.

François, M. 1928 *C.R. Soc. Biol. Paris*, **98**, 152.

Gemant, A. 1936 *Physics*, **7**, 311.

Gemant, A. 1944 *Phil. Mag.* **25**, 540.

Grunberg, L. & Nissan, A. H. 1945 *Nature*, **156**, 241.

Hoagland, H. 1935 *Pace-makers in relation to aspects of behavior*. New York: Macmillan.

Love, A. E. H. 1906 *A treatise on the mathematical theory of elasticity*, 2nd ed. Camb. Univ. Press.

Mooney, M. 1940 *J. Appl. Phys.* **11**, 582.

Noüy, L. du 1936 *Biological time*. London: Methuen.

Nutting, P. G. 1921*a* *J. Franklin Inst.* **191**, 679.

Nutting, P. G. 1921*b* *Proc. Amer. Soc. Test. Mater.* **21**, 1162.

Piéron, H. 1945 Presidential Address to French Ass. for Advancement of Science, Paris.

Poincaré, H. 1904 *La Valeur de la Science*, p. 44. Paris: Flammarion.

Prache, P. 1930 *Caoutch. et Gutta-p.* **32**, 15003.

Reiner, M. 1933 *Naturwissenschaften*, **21**, 294.

Schrödinger, E. 1944 *What is life?* Camb. Univ. Press.

Whitehead, J. B. 1935 *Impregnated paper insulation*, p. 192. New York: Wiley.

Kinetic boundary friction

By J. R. Bristow, *Engineering Division, National Physical Laboratory, Teddington, Middlesex*

(*Communicated by Sir Charles Darwin, F.R.S.—Received* 10 *October* 1945—*Read* 28 *February* 1946)

[Plate 3]

It is shown that the exhibition of smooth sliding and relaxation oscillations, or 'stick-slips', under conditions of boundary lubrication, when frictional forces are measured by the deflexion of an elastic system, may be explained as due to the dependence of kinetic friction on velocity. In the cases giving smooth sliding, kinetic friction decreases as velocity decreases, at very low speeds; for the cases giving relaxation oscillations kinetic friction increases as velocity decreases. That is, sliding under boundary conditions is not inherently discontinuous, any discontinuous motion being due to the dynamics of the measuring instrument, and is the result of kinetic friction increasing as velocity decreases. Curves of boundary friction against velocity, using various sliding surfaces, have been determined for a number of lubricants, which show both the above-mentioned types of friction-velocity relationship; and the dependence is shown of kinetic boundary friction on molecular weight for a series of esters of the fatty acids, on percentage of fatty oil in a compounded lubricant (actually oleic acid in mineral oil) and on temperature for a pure substance and a mineral oil. The measurements with the series of esters show some agreement with results given by Fogg (1940). The mixtures of oleic acid with mineral oil give decreasing kinetic friction with increasing percentage of oleic acid right up to 100 % oleic acid. The effect of temperature on the dependence of friction in velocity shows that the temperature at which relaxation oscillations first occur depends on the speed of sliding, from which it appears that measurements of the temperature at which relaxation oscillations start at constant sliding speed (Frewing 1942) are not a measure of the temperature at which there is a discontinuity in the properties of the boundary layer.

INTRODUCTION

Although numerous investigations into the nature of boundary friction, and friction in general, have been made over a period of some 250 years, it is only during the present century that any real advance has been made towards some slight understanding of the phenomenon of boundary lubrication and the nature of boundary lubricating layers.

Now, of necessity, investigations into the nature of friction under conditions of boundary lubrication consist mainly of measuring coefficients of friction. Examinations of boundary layers by electron diffraction, etc., are only of a subsidiary nature. Hence the investigator of boundary-friction phenomena is in the somewhat difficult position of trying to obtain fundamental information by means of something (friction) the mechanism of which is not understood. All the investigator can hope to do is to interpret the values of coefficients of friction obtained in the light of information available on the physical and chemical properties of thin films of organic substances (lubricants are almost invariably organic) on solid, usually metal, surfaces. That useful literature of the desired nature is almost non-existent is well

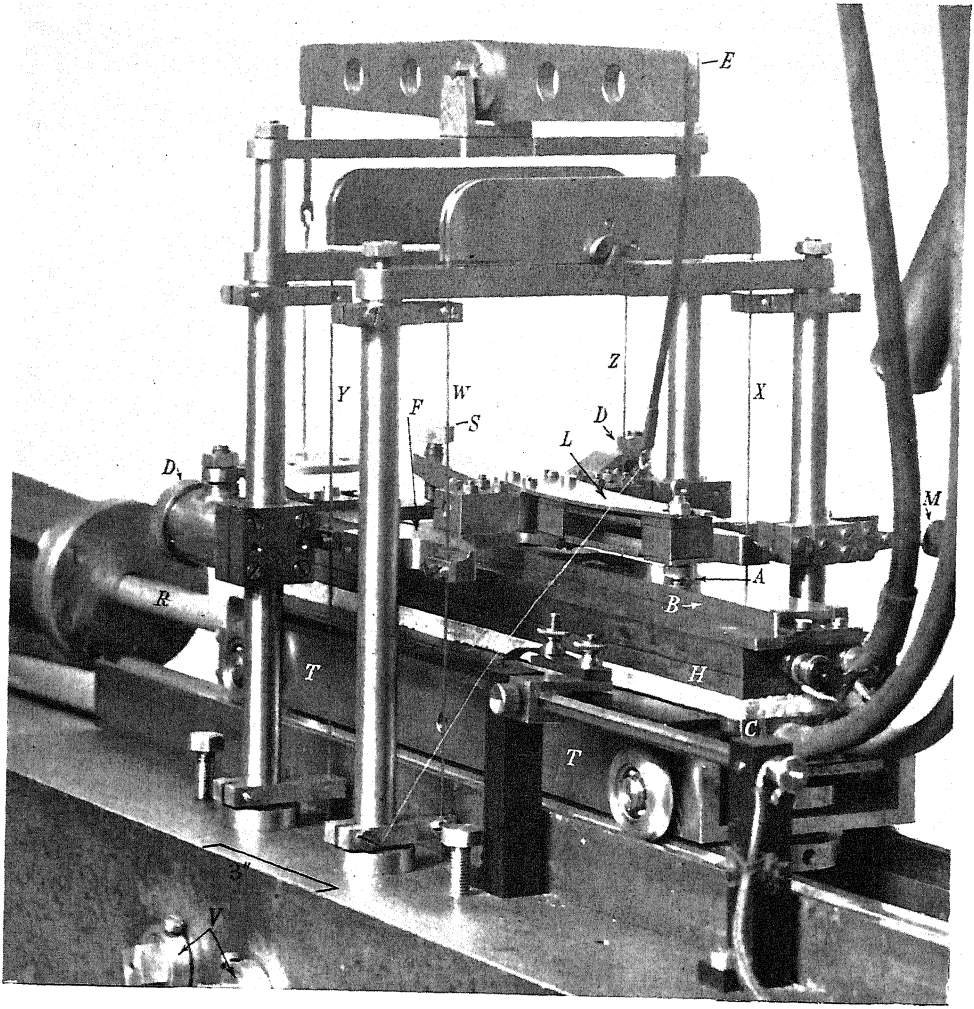


FIGURE 3

known. However, the nature of thin films of organic substances on water has been very fully investigated since Langmuir's paper on the subject in 1917 (see Adam 1938), and this knowledge was drawn on by Hardy (1936) to interpret his friction measurements. To explain the continuous change in frictional properties with chain length for series of long-chain compounds, fatty acids, alcohols, etc., Hardy assumed that the power to act as a boundary lubricant is due to an adsorbed layer which is monomolecular and orientated, as are films of the same substances as condensed films on water. Orientation of lubricant layers on metal surfaces has been confirmed by X-ray and electron-diffraction experiments by many investigators including Trillat (1925), Bragg (1925) and Clark, Sterret & Lincoln (1936). An attempt to take the analogy between films on water and film on metal surfaces still further has been made by Frewing (1942). It is presumed by Frewing that his friction measurements determine the temperature at which the orientated boundary layer becomes disordered and that surface films may change from condensed to expanded types; but it is thought that the experimental evidence is insufficient to warrant all the conclusions drawn.

However, full use of the available variables, which might reasonably be expected to affect boundary friction, has not been made until quite recently; in particular, the effects of speed and temperature. For example, the greatest number of friction measurements made is probably of values of 'static' friction. In all Hardy's work values of the coefficient of 'static' friction were determined, the static value being that at which motion is first apparent with slowly increasing tangential load. It was, however, shown, as long ago as 1877, by Kimball that kinetic friction (F_k) may decrease with decreasing velocity at very low speeds, and Kaufmann (1910) and Jacobs (1912) suggested that kinetic friction tends to zero as the speed falls to zero and that at high speed the friction approaches an asymptotic value (F_a) (see figure 1*c*). Hence, for any value of tangential force applied less than F_a uniform motion will ensue. This is as Hardy found for the longest chain compounds he used, e.g. palmitic acid, but he failed to appreciate that 'static' friction (F_S) might in fact not exist, for clearly the value of F_S will depend on the sensitivity of the method used to ascertain when motion has started, unless $F_S > F_K$, when it would seem that no ambiguity could occur. It will be seen later, however, that even if F_S is apparently greater than F_K unambiguous determinations of 'static' friction are not necessarily possible.

If the frictional force is measured with an elastic system, which carries one friction surface and presses it against a moving surface, the elastic system may perform relaxation oscillations, called in the literature 'stick-slips' (see Bowden & Leben 1939, 1940; and Muskat, Morgan & Reed 1941). The term relaxation oscillation is preferred to 'stick-slip', since 'stick-slip' appears to be associated with 'fluctuations in friction' rather than with the motion of an elastic system. At first an explanation of these relaxation oscillations—and also of the cases where no oscillations occur—was sought in the difference between 'static' and kinetic friction and the dependence of kinetic friction on velocity. In actual fact the relation of 'static' friction (if it exists) to the work discussed later proves to be obscure, and simple considerations of the dynamics of the apparatus used and examination of the results obtained indicates that, if relaxation oscillations occur, it is sufficient that kinetic friction increases with decreasing velocity, at low speeds (see figure 1*b*), and, if no relaxation oscillations occur, that kinetic friction decreases with decreasing velocity, at very low speeds (figure 1*c*). The presence or absence of 'stick-slips' at any particular temperature is therefore dependent on the speed of sliding, the dynamics of the instrument and the nature of the dependence of kinetic friction on velocity. The work of Bowden & Leben (1939, 1940) and Muskat *et al.* (1941), and the work of Frewing (1942) on the effect of temperature on boundary lubrication all suffer from the failure to consider the dynamics of the instrument used.

In the following pages elementary considerations of the behaviour of the friction measuring instrument used, and several representative examples of kinetic boundary-friction determinations are given. The discussion also covers the apparatus used by Muskat *et al.* (1941), Bowden & Leben (1939, 1940) and Frewing (1942). It is considered that as yet knowledge of the phenomena exhibited under boundary conditions is not sufficient to make it possible to deduce from the results obtained a reasonable theory of either kinetic boundary friction or the nature of the molecular layers responsible for boundary lubrication. Hence the results are given quite separately and discussed in connexion with comparable results of other workers and not in relation to one another.

ELEMENTARY THEORY OF THE BEHAVIOUR OF AN ELASTIC SYSTEM DEFLECTED BY BOUNDARY-FRICTION FORCES AT A MOVING SURFACE, AND PERTINENT EXPERIMENTS

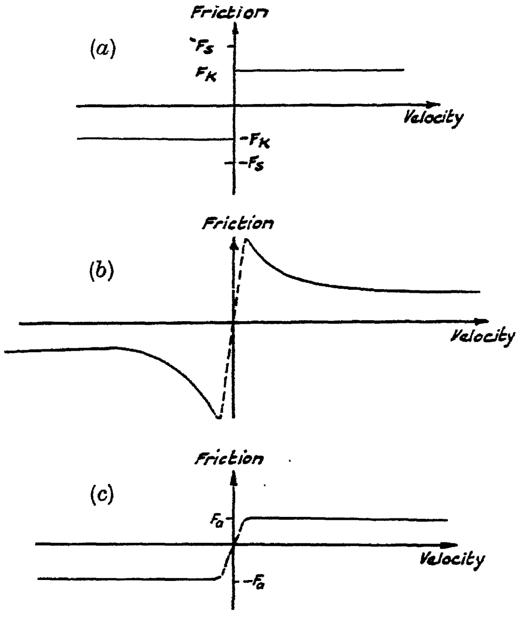
The considerations put forward here concern the motions of an elastically restrained upper surface which are produced by the frictional forces at its point of contact with a lower moving surface against which it is pressed. The simple assumptions made in the first instance are that static friction (F_S) is greater than kinetic friction (F_K), and that kinetic friction is independent of velocity (see figure 1*a*). This case has been considered in some detail by Thomas (1930), and it is easy to show that the elastic system exhibits relaxation oscillations for slow sliding speeds.

Consider the upper friction surface to be in its undeflected position; then, as the lower surface moves it carries the upper with it until the spring force (F) (plus a small force due to viscous damping) exceeds the static friction force and slip takes place. The system will now perform oscillations according to the equation, with suitable choice of units,

$$\ddot{x}+k\dot{x}+x = \pm F_K.$$

If F_K is not much less than F_S only one swing is made about $+F_K$ before the system comes to rest relative to the moving surface in such a position that F lies between $+F_K$ and $-F_K$ and oscillations must cease. Then the upper surface is carried forward again by the lower until F_S is again exceeded and the cycle repeated.

This is illustrated in figure 2 for various amounts of damping. If the damping is such as to make the system aperiodic, after the first deflexion and slip, the deflexion will become constant at F_K.



FIGURES 1 *a*, *b*, *c*

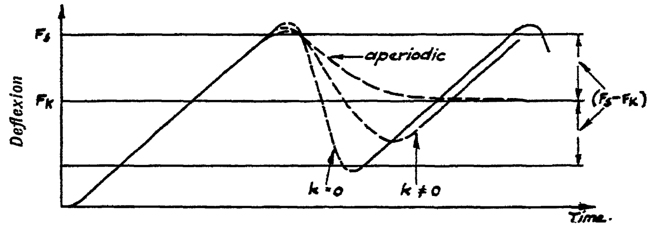


FIGURE 2

If $F_K \geqslant F_S$ clearly no oscillations occur, the deflexion being stable at F_K.

If the system is virtually undamped, the natural period of the elastic system short and the velocity of the moving surface low, the relaxation oscillation will be 'saw-toothed' in shape, as are the 'stick-slip' figures given by Bowden & Leben (1939). That the motion during slip is approximately sinusoidal has been shown by Bristow (1942) and Muskat *et al.* (1941).* If the velocity during 'slip' is sufficiently high a noticeable temperature 'flash' may be produced by the work done against the frictional force. Calculations of the rise of temperature to be expected on the lines given by Blok (1937) and Jaeger (1942) give values of the same order of magnitude as those measured by using sliding contacts of dissimilar metals as thermocouples. The highest temperature flash that has been recorded on an instrument identical with that of Bowden & Leben is about 30° C. It would appear that the observations published by Bowden & Leben (1939) are sufficiently explained by the above considerations.

---

* A case shown by Muskat *et al.* (1941) (figure 7, their paper) where F_K is small compared with F_S shows two 'swings' between each slip about first $+F_K$ and then about $-F_K$, indicating that F_S does not come into play immediately because the relative velocity between the surfaces is zero, and that the condition of 'stick' is that F falls between $+F_K$ and $-F_K$.

However, further observations of the motions of such an elastic system indicated that the above simple assumptions regarding boundary friction were only a first approximation.

The important observations were: (*a*) if the system is damped to be aperiodic, oscillations did not cease; (*b*) with a virtually undamped system the size of the relaxation oscillations decreases as the speed of the moving surface increases, eventually becoming vanishingly small; at still greater speeds quasi-sinusoidal oscillations may be maintained in the system; and (*c*) the total movement of the elastic system during 'stick' was less than the movement of the lower surface in the same time, and decreased with increase of lower surface speed to become necessarily zero when oscillations ceased. From these experimental observations it is clear that F_K is not independent of velocity, and it appears that, in fact, F_K decreases rapidly as velocity increases, quickly becoming sensibly constant (see full-line figure 1*b*). It may readily be shown that if F_K varies as in figure 1*b*, the elastic system may exhibit relaxation oscillations depending on the amount of viscous damping.

Let the frictional force be $F(v)$, where v is the relative velocity between the two surfaces. Then the equation of motion, when the lower surface is moving with velocity V, with suitable choice of units, is

$$\ddot{x}+k\dot{x}+x = F(V-\dot{x}). \tag{1}$$

Suppose the motion is steady, then x is constant, i.e. $x = F(V)$. This condition may be unstable. Suppose a small perturbation given by $x = F(V)+y$, then substituting in (1)

$$\ddot{y}+k\dot{y}+F(V)+y = F(V-\dot{y}).$$

Expanding the right-hand side by Taylor's theorem

$$\ddot{y}+k\dot{y}+y = -F(V)+[F(V)-\dot{y}F'(V)+\ldots],$$

neglecting terms in the expansion after the second

$$\ddot{y}+(k+F'(V))\dot{y}+y = 0.$$

For F_K varying with velocity as in figure 1*b*, $F'(V)<0$. Hence if k is negligible the motion is certainly unstable, but stability may be produced if the value of k is sufficiently great. If F_K varies as in figure 1*c* obviously the motion is always stable. (Case of Jacobs 1912 and Kaufmann 1910.)

These considerations and experimental observations agree with the work of Papenhuyzen (1938) on the sliding of rubber on glass and other surfaces. Papenhuyzen was also able to show, however, that on reducing the speed of the lower friction surface sufficiently, oscillations ceased and smooth sliding ensued at values of friction (F_K) which decreased with speed. This is indicated by the dotted extensions to the curves in figure 1*b*.

In concluding this section it may be emphasized that on the foregoing analysis relaxation oscillations (or 'stick-slips') are fundamental to the type of instrument, having regard to the dependence of friction on velocity and the dynamics of the elastic system, and not that 'these discontinuities are inherent in the sliding of metals' (Bowden & Leben 1939). Hence it is not possible to deduce from the papers of Bowden & Leben (1939) or others describing work with similar apparatus any information regarding kinetic friction other than that $F_K < F_S$, if 'stick-slips' occur.

EXPERIMENTAL

The friction-measuring apparatus was designed to determine curves of boundary friction against velocity at low speeds on the basis of the foregoing analysis, which indicates that to determine F_K the period of the elastic measuring system is not of primary importance but that high damping is required to make the system stable.

A general view of the instrument is shown in figure 3, plate 3, and the essential part of the system diagrammatically in figure 4. The tangential forces due to friction are measured between the upper spherically ended surface, *A*, and the lower flat surface, *B*. The use of a spherically ended surface on a flat in order to obtain boundary conditions is as described by Hardy (1925). The flat surface is carried on a trolley, *T*, driven at various constant speeds by hydraulic ram designed to impart as smooth a motion as possible to the trolley. The driving rod, *R*, of the ram, is just visible in figure 3. The speed is controlled by the rate of flow of the hydraulic fluid into the system. In order to work at temperatures above normal room temperatures, a heater, *H*, is built into the trolley; the cooling system, *C*, beneath the heater is to keep the heat localized to the friction surface. The trolley wheels run on horizontal rails, one of which is a V-groove to define the direction of motion. The speed of the lower surface is measured by timing over a fixed distance by a stop-watch electrically operated by the trolley. The top spherically ended surface is carried on the rigid frame, *F*, attached to the four wires *W*, *X*, *Y* and *Z*, which are arranged to be in equal tension in pairs (i.e. $T_W = T_X$, $T_Y = T_Z$) and are tensioned by the spindles, *V*. Then, with the top surface fixed on the centre line of the frame, *F*, it will be deflected parallel to the direction of motion of the trolley by the forces at its point of contact with the lower surface. This arrangement was adopted, as it was at one time thought that twisting of the surfaces relative to each other might produce increased adhesion between the surfaces (Bristow 1942). This was later found to be an unnecessary refinement. The elastic system is heavily damped by the arrangements *D*, *D*, each of which consists of a piston attached to the carriage and moving in a cylinder, carried on the body of the apparatus. There is a small clearance between piston and cylinder which is filled with oil. Deflexions of this elastic system are measured by a mirror-lamp system *M*, and are observed visually on a scale (not shown). The deflexions are calibrated in terms of force by pulling the system forwards by a string attached to the upper friction surface which runs over a pulley (not shown) to a scale pan. Load is applied through the cantilever, *L*, by turning the screw, *S*, which bears on the frame, *F*. The load is determined by measuring the force necessary to lift the top surface out of contact with the lower by means of the weigh beam, *E*. The observations taken are deflexion and time to travel a certain distance, from which the coefficient of friction and velocity can be determined, and curves of friction against velocity constructed. In some cases, with the heaviest convenient damping, small oscillations occur at the very lowest speeds. For such cases the curves through the experimental points are drawn through the mean of the points obtained from the maximum and minimum deflexions during the oscillations. This does not produce any great uncertainty in the curves, as in such cases the coefficient of friction is always rising very rapidly as velocity decreases.

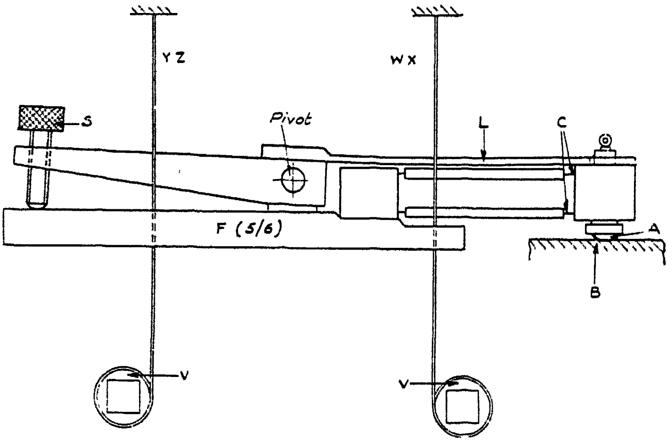


FIGURE 4. Kinetic friction apparatus. Details of suspension.

Two methods of cleaning the friction surfaces have been used: (1) polishing with 0000 emery paper, scrubbing in strong caustic soda solution with cotton-wool until the entire surface is completely wetted with water, washing in water and drying with acetone; and (2) cutting the surface with a clean tool as described by Ernst & Merchant (1940). In general the first method has been used on steel surfaces and the second on soft materials. All the results given were obtained under conditions of constant load and with excess lubricant present.

RESULTS

A selection of curves of coefficient of friction against velocity are given here which are directly comparable with other published work on boundary friction, and which illustrate the discussion already given. It will be seen that these results do not conflict with those given by other workers, but do add information which is not in agreement with the interpretations given by previous investigators.

Results are given for:

(i) A series of ethyl esters of saturated fatty acids for two pairs of sliding surfaces, (1) hard steel on hard steel, and (2) phosphor bronze on hard steel, at room temperature.

(ii) A mineral oil containing additions of various percentages of oleic acid, with sliding surfaces of (1) steel on steel, and (2) brass on hard steel, at room temperature.

(iii) For (1) ethyl palmitate lubricating the sliding of tin on hard steel and for (2) a mineral oil used as lubricant for steel sliding on hard steel, at various temperatures.

(i) *The kinetic boundary friction of a series of esters*

The series of esters used is given in table 1, and the measurements were made at room temperature. The results of the determinations are given in figures 5 and 6, the curves being numbered for the corresponding lubricants as in table 1. The nature of the results may perhaps be seen better in the plot of coefficient of friction against molecular weight given in figures 7 and 8.

TABLE 1. ESTERS USED

number	ester	formula	molecular weight
1	ethyl formate	$H.CO.O.C_2H_5$	74
2	ethyl acetate	$CH_3.CO.O.C_2H_5$	88
3	ethyl propionate	$C_2H_5.CO.O.C_2H_5$	102
4	ethyl valerate	$C_4H_9.CO.O.C_2H_5$	130
5	ethyl oenanthate	$C_6H_{13}.CO.O.C_2H_5$	158
6	ethyl capryllate	$C_7H_{15}.CO.O.C_2H_5$	172
7	ethyl pelargonate	$C_8H_{17}.CO.O.C_2H_5$	186
8	ethyl laurate	$C_{11}H_{23}.CO.O.C_2H_5$	228
9	ethyl palmitate	$C_{15}H_{31}.CO.O.C_2H_5$	284

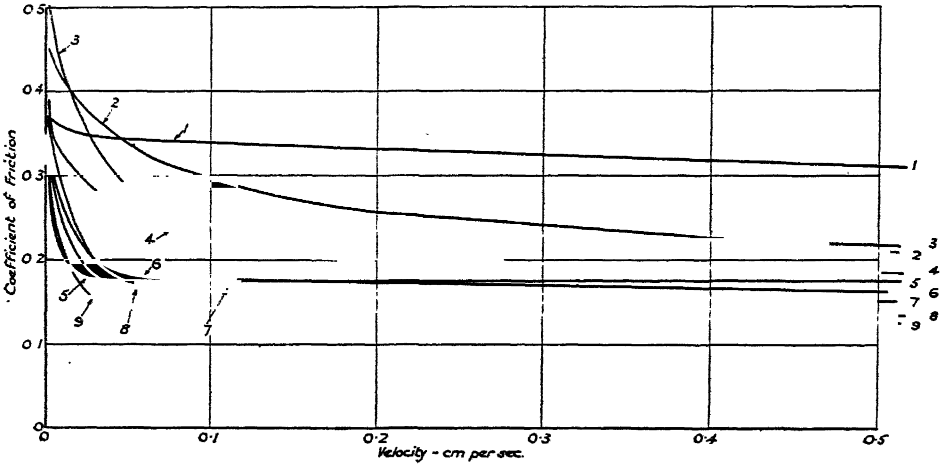


FIGURE 5. Hard steel sliding on hard steel lubricated with ethyl esters of saturated fatty acids at room temperature.

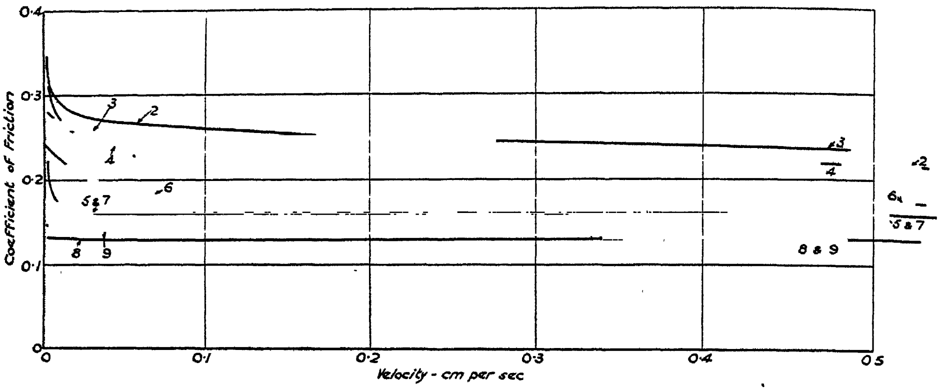


FIGURE 6. Phosphor bronze sliding on hard steel lubricated with ethyl esters of saturated fatty acids at room temperature.

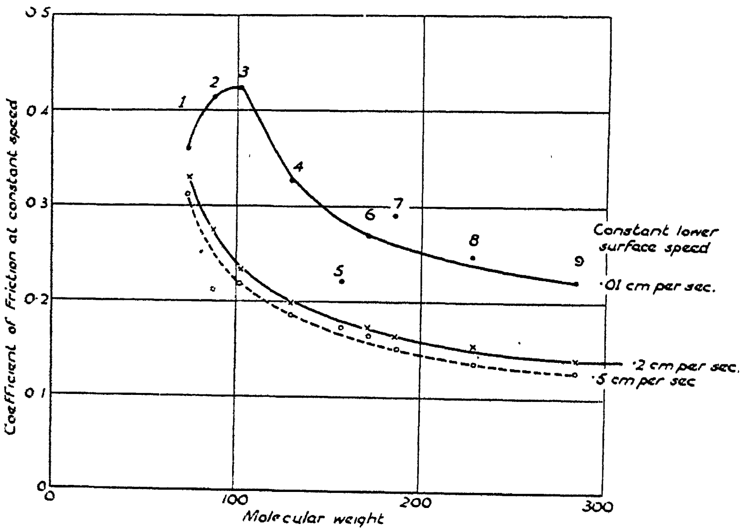


FIGURE 7. Hard steel sliding on hard steel lubricated with ethyl esters of saturated fatty acids at room temperature.

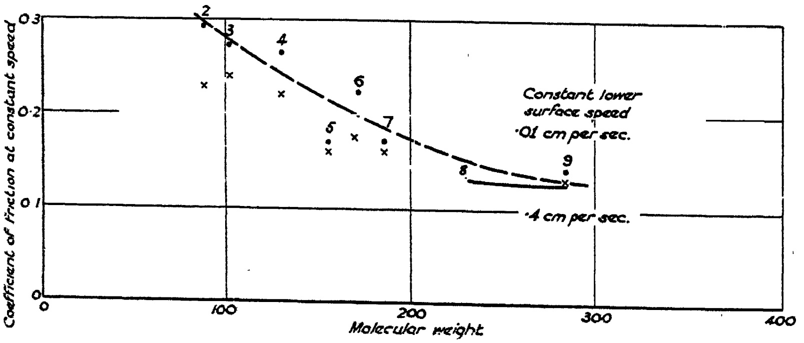


FIGURE 8. Phosphor bronze sliding on hard steel lubricated with ethyl esters of saturated fatty acids at room temperature.

The results given in figure 7 for hard steel on hard steel are comparable with those of Fogg (1940) for 'static' friction measurements on the same series of esters, using a modified Deeley machine. Fogg's measurements gave a curve of coefficient of 'static' friction against molecular weight similar to the one given for a constant velocity of 0·01 cm./sec., with the maximum of the curve between ethyl propionate (3) and ethyl valerate (4). Roughly the explanation given, on the assumption that the coefficient of friction is inversely proportional to the distance between the two friction surfaces as determined by a monomolecular adsorbed layer on each surface, was that the esters are attached to the surface by the double-bond oxygen atom in the molecule with the two branches (oxygen atom and ethyl group, and varying fatty acid chain) standing at angles determined by their relative weights, the greater the difference in branch size the more perpendicular the longer branch stands to the surface. Hence a minimum thickness of adsorbed film may be expected between ethyl propionate (3) and ethyl valerate (4), at which the maximum friction of the series was found.

However, for the determinations of kinetic boundary friction a maximum is only found at slow speeds for hard steel on hard steel. At high speeds, the friction falls continuously with increasing molecular weight for hard steel sliding on hard steel, and for phosphor bronze on hard steel no maximum is shown even at very low speeds (in this latter case the points are somewhat scattered).

It is apparent that the explanation given by Fogg is insufficient to explain all of the above results.

(ii) *The kinetic boundary friction of a mineral oil containing additions of oleic acid*

The results are given in figures 9 and 10. It will be seen that the coefficients of kinetic boundary friction are continuously reduced by increasing additions of oleic acid to the mineral oil, the oleic acid having good properties as a boundary lubricant, and the mineral oil poor properties. The effect of additions of fatty oil is greatest at low concentrations, and it will be noted that at high concentrations of fatty substance the friction-velocity curve changes in type from that shown in figure 1*b* to that shown in figure 1*c*, for brass on steel, but not for steel on steel. It is clear that compounding a mineral oil with a fatty oil produces considerable improvements in boundary lubricating properties even for small additions, say up to 5 % fatty oil, but that the friction still continues to decrease slowly right up to 100 % of the fatty constituent. At high speeds the effect of additions of the fatty substance tends to decrease.

Such a result is to be expected, since the boundary layer will probably always contain a greater percentage of the more 'active' constituent than the solution. But apparently even at high concentrations of the 'active' substance (oleic acid) the boundary layer still contains an appreciable amount of the less active substance (mineral oil). The effects to be expected will, of course, depend on the relative surface 'activities' of the constituents of any particular mixture.

In figures 11 and 12 coefficients of friction are plotted against concentration of oleic acid in mineral oil at certain constant speeds, and show very clearly how small amounts of oleic acid reduce the friction very considerably. These graphs may be compared with those given by Fogg & Hunwicks (1940) for additions of rape oil to

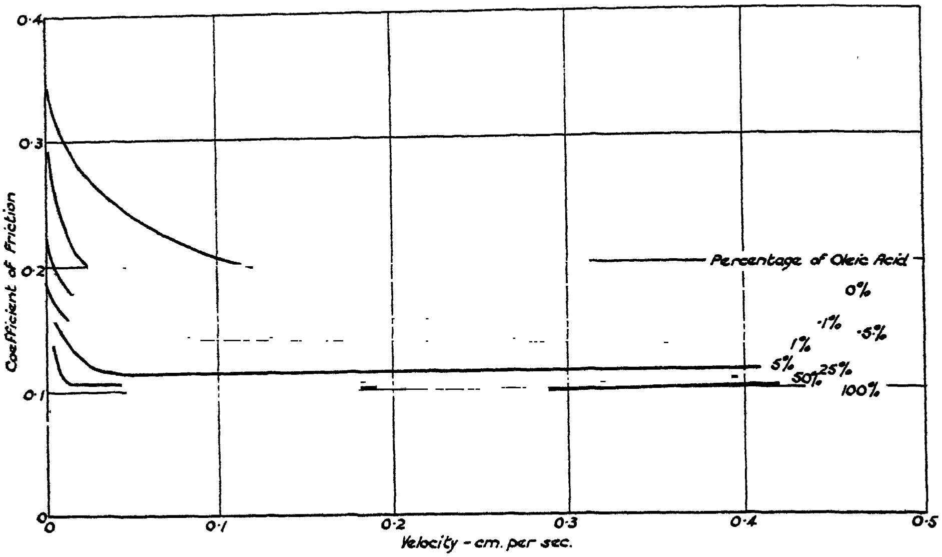


FIGURE 9. Brass sliding on steel lubricated with mineral oil—oleic acid mixtures at room temperature.

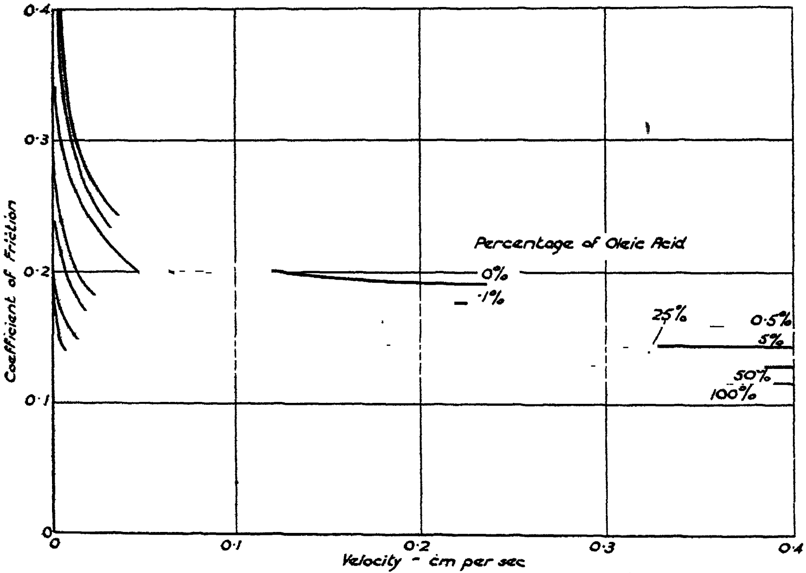


FIGURE 10. Steel sliding on steel lubricated with mineral oil—oleic acid mixtures at room temperature.

a mineral oil and oleic acid to B.P. paraffin, for determination of 'static' friction with a modified Deeley machine. The curves given by Fogg & Hunwicks are the same shape, and show the same facts as those given in this paper for low velocity.

The effect shown of the decreasing effect of additions of oleic acid at higher speeds has not so far been satisfactorily explained.

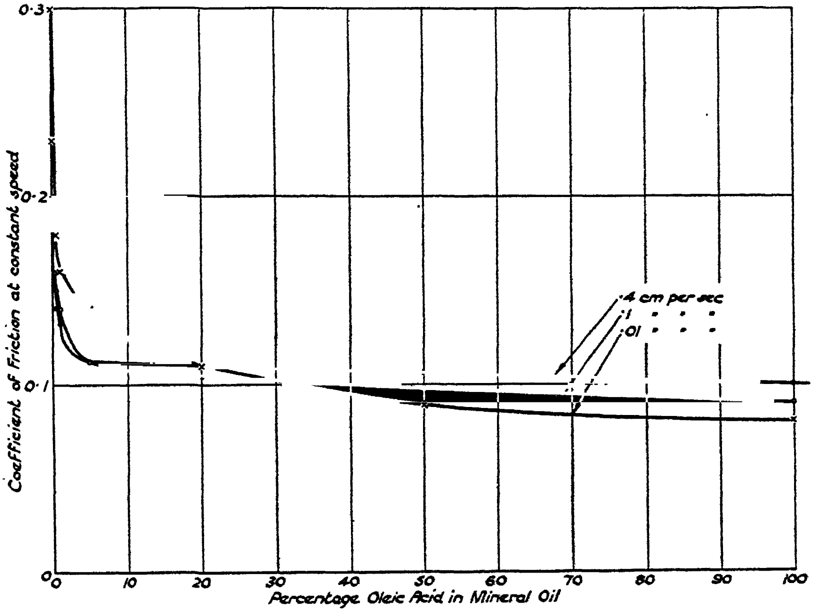


FIGURE 11. Brass sliding on steel lubricated with mineral oil—oleic acid mixtures at room temperature.

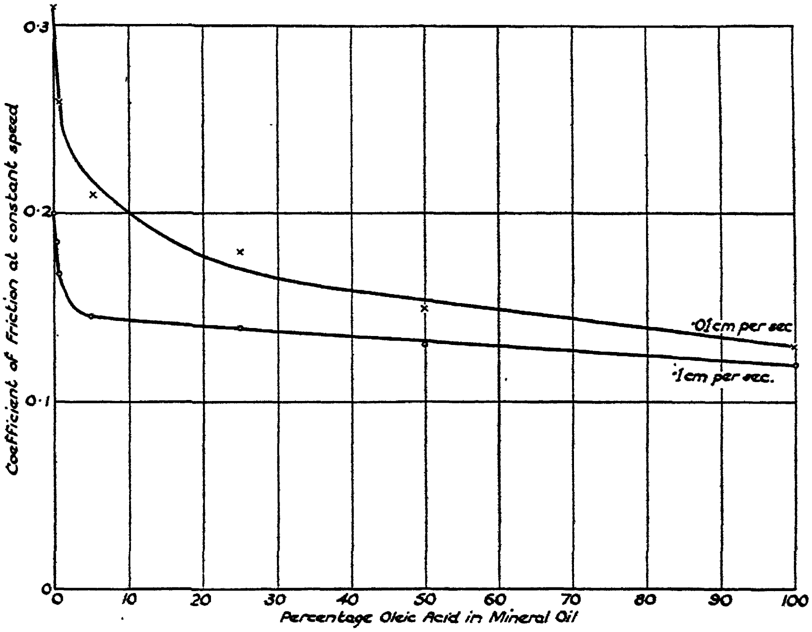


FIGURE 12. Steel sliding on steel lubricated with mineral oil—oleic acid mixtures at room temperature.

(iii) *The effect of temperature on kinetic boundary friction*

The effect of temperature on kinetic boundary friction for various surfaces and lubricants appears to be somewhat complicated, and the discussion here is confined

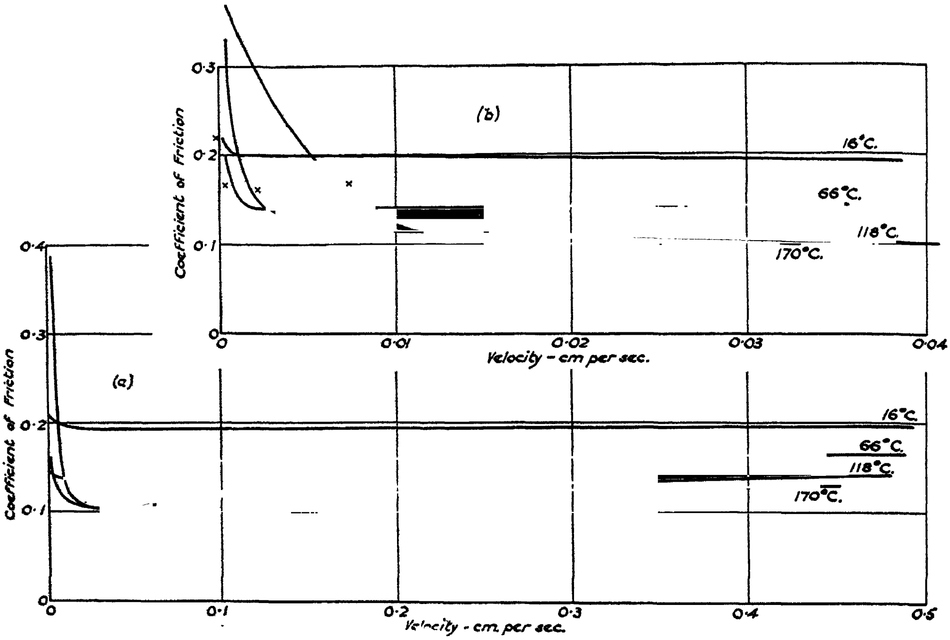


FIGURE 13. Tin sliding on hard steel lubricated with ethyl palmitate at various temperatures.

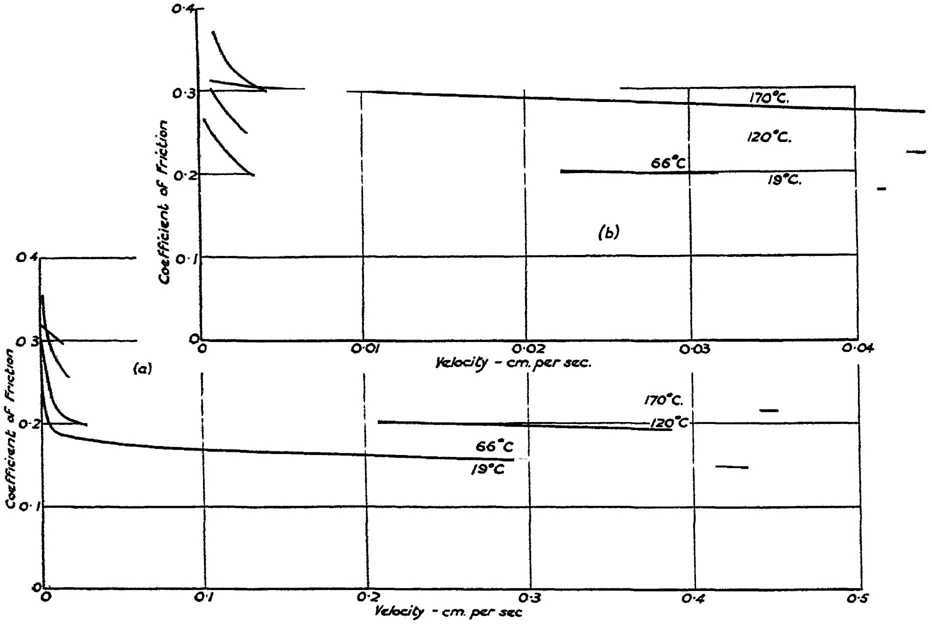


FIGURE 14. Steel sliding on hard steel lubricated with a mineral oil at various temperatures.

to two of the simpler cases investigated, namely, (1) tin sliding on hard steel lubricated with ethyl palmitate, and (2) steel sliding on hard steel lubricated with a thin mineral oil. The friction velocity curves for several temperatures are given in figures 13*a* and *b* and 14 *a* and *b*. Figures 13*b* and 14*b* give the low-velocity ends of figures 13*a* and 14*a* in greater detail by plotting velocity on a larger scale.

(1) *Ethyl palmitate lubricating tin on hard steel.* This case is of particular interest in relation to the work published by Tabor (1940, 1941), Frewing (1942) and Hughes & Whittingham (1941).

At the higher velocities the coefficient of kinetic friction decreases with increase of temperature; at the lower velocities, however, the coefficient of kinetic friction increases, the increase extending to greater speeds as the surfaces get hotter. The change of shape of the curves of friction against velocity shown at very low velocities necessarily means that the tendency to give relaxation oscillations increases as the temperature is raised. The crosses on the graphs of figure 13 indicate approximately the highest velocity at which relaxation oscillations are first apparent. The velocity at which the oscillations vanish is, of course, dependent on the amount of damping employed. It is clear, therefore, that if the lower surface velocity in the apparatus is constant at, say, 0·01 cm./sec., no relaxation oscillations will occur even if the temperature is raised to 170° C. On the other hand, if the surface velocity is less than about 0·002 cm./sec., relaxation oscillations will occur at room temperature. It is clear that over this range of temperature 'stick-slips' can start at any desired temperature as the surface temperature is raised by choosing the correct velocity of the lower surface. Using steel on steel lubricated with ethyl palmitate the change in shape of the curves at very low velocity is essentially the same as for the above case.

That some change is taking place, as the temperature is raised, which is connected with the nature of the molecule of the lubricant and surfaces concerned, cannot be doubted. But that this change is necessarily connected with the sudden disorientation or desorption of an adsorbed film, and that the temperature is characteristic of the lubricant (i.e. that there is some sudden change at this temperature) as suggested by Tabor (1940, 1941) and supported by Frewing (1942), cannot be maintained. In this connexion it may be noted that the values given by Hughes & Whittingham for the temperature at which 'stick-slips' start, with fatty acids as lubricants, are between 20 and 50° C higher than the values given by Frewing.

(2) *A mineral oil lubricating steel on hard steel.* For this case the coefficient of friction increases with rise of temperature over the whole of the velocity range used (except for 170° C at very low speed). Here again, however, the tendency to give relaxation oscillations ('stick-slips') increases as the temperature rises. As in (1) above it is possible to make the temperature at which relaxation oscillations start any desired value by fixing correctly the velocity of sliding.

At this stage, it is not possible to give a satisfactory explanation of these phenomena.

This paper is published by permission of the Director of the National Physical Laboratory and the Director General of Scientific Research and Development, Ministry of Supply, for whom the investigation was made.

REFERENCES

Adam 1938 *The physics and chemistry of surfaces.* Oxford University Press.
Blok 1937 *Inst. Mech. Engrs., Discussion on Lubrication and Lubricants, Group IV*, p. 26.
Bowden & Leben 1939 *Proc. Roy. Soc.* A, **169**, 391.
Bowden & Leben 1940 *Phil. Trans.* A, **239**, 1.
Bragg 1925 *Proc. Roy. Inst.* **24**, 481.
Bristow 1942 *Nature*, **149**, 169.
Clark, Sterrett & Lincoln 1936 *Industr. Engng Chem.* **28**, 1318.
Ernst & Merchant 1940 *Mass. Inst. Tech., Proc. Conf. on Friction and Surface Finish*, p. 76.
Fogg 1940 *Proc. Phys. Soc.* **52**, 239.
Fogg & Hunwicks 1940 *J. Inst. Petrol*, **26**, 1.
Frewing 1942 *Proc. Roy. Soc.* A, **181**, 23.
Hardy 1936 *Collected Works.* Cambridge University Press.
Hardy 1925 *Proc. Roy. Soc.* A, **108**, 2.
Hughes & Whittingham 1941 *Trans. Faraday Soc.* **38**, part 1, 9.
Jacobs 1912 *Ann. Phys., Lpz.*, **38**, 126.
Jaeger 1942 *J. Roy. Soc.* N.S.W. **56**, 203.
Kaufmann 1910 *Phys. Z.* **11**, 985.
Kimball 1877 *Amer. J. Sci.* p. 353.
Langmuir 1917 *J. Amer. Chem. Soc.* **39**, 1848.
Muskat, Morgan & Reed 1941 *J. Appl. Phys.* **12**, 743.
Papenhuyzen 1938 *Ingenieur, Utrecht*, **53**, 75.
Tabor 1940 *Nature*, **145**, 308.
Tabor 1941 *Nature*, **147**, 609.
Thomas 1930 *Phil. Mag.* **9**, 329.
Trillat 1925 *C.R. Acad. Sci., Paris*, **180**, 280.

A general kinetic theory of liquids.
II. Equilibrium properties.

By H. S. Green, *Edinburgh University*

(*Communicated by M. Born, F.R.S.—Received* 4 *June* 1946)

The object of this paper is to expand and develop the equilibrium theory outlined in the first part of this series, within the framework of a general kinetic theory of fluids.

The entropy of the fluid is first defined in such a way that agreement with the definition of statistical mechanics is obtained in equilibrium conditions. It is then found possible, provided care is taken to avoid certain well-known fallacies in the application of statistics to mechanical systems, to prove that the quantity defined will in general increase with time towards a certain maximum. The equilibrium solution of the fundamental equations is thereby derived without any appeal to statistical mechanics. At the same time a direct proof is offered of certain general formulae in statistical thermodynamics.

The second half of the paper is devoted to a more detailed study of the equilibrium state by a method proposed in part I of the series. It is shown that the liquid state is distinguished from the gaseous state by the existence of real roots of a certain transcendental equation. Approximate but tractable expressions are found for all the virial coefficients in the gaseous phase. An equation of state is derived with two branches, which are identified with the gaseous and liquid states respectively.

1. INTRODUCTION

In part I of this series (Born & Green 1946), Professor Max Born and the author gave a general survey of a theory designed to describe both equilibrium and dynamical properties of liquids. A brief treatment of the equilibrium state was included, sufficient to show that results consistent with the application of statistical mechanics could be obtained. There were, however, two important gaps in our outline, which it is the purpose of the present paper to fill.

In the first place, we found it necessary to appeal to statistical mechanics in order to prove that the Maxwellian distribution of velocities in a molecular assembly was a *unique* consequence of our equilibrium theory, though we were afterwards able to show that by assuming this, the complete statistical mechanical solution could be obtained. It is obviously desirable, however, to make an exposition of this kind independent, as far as possible, of external theories, even though their subject-matter may overlap to a certain extent with our own. The first part of this paper is therefore devoted to a generalization of the famous Boltzmann H-theorem (Chapman & Cowling 1939), which enables us to establish our treatment of the equilibrium properties on an independent basis. At the same time a direct proof is offered of certain general formulae of statistical thermodynamics.

Secondly, although the object of this series is specifically to provide a theory of the liquid state, in part I no indication was given of the role played by this particular phase in our theory. Indeed, in its present form, it is equally applicable to the gaseous phase, and, apart from modifications probably of a quantum-mechanical nature, to the solid state as well. It just happens that there are already theories which adequately describe these other states. However, some explanation of the phenomenon of condensation, and the distinction between the gaseous and liquid states, is plainly required, and this it is the endeavour of the second part of the present paper to give.

A theory of condensation, based on statistical mechanics, has already been elaborated by Mayer and others (Mayer 1937; Born & Fuchs 1938; Kahn 1938), who showed that condensation is associated with the divergence of the series obtained by expanding the normalized phase integral in powers of the density. The nature of this singularity, and of the continuation beyond, has, however, so far remained somewhat obscure. The investigation which follows indicates that the singularity is, in fact, a branch point, from which spring two branches associated with the gaseous and liquid states respectively.* With the aid of an approximate solution of an integral equation derived in part I, it has been found possible to derive equations of state for both phases, and any other equilibrium properties which may be required. This part of the work offers many opportunities for numerical comparison with experiment, but it has not seemed worth while to specialize in this way at present; on account of the approximation involved, qualitative rather than exact quantitative agreement may be expected. However, the equilibrium theory now seems sufficiently well established to enable us to proceed to the more interesting dynamical properties, some of which Professor Born and the author hope to treat in the third part of this series.

2. PRELIMINARY

A fundamental theorem in the kinetic theory of gases, due to Boltzmann, shows that a certain quantity H, closely connected with the entropy S, cannot, on the average, increase, but must tend to a fixed minimum, associated with the equilibrium state. Both the definition of H, and the proof of the so-called H-theorem, must, however, be generalized to apply to the liquid phase.

The appropriate definition is suggested by known results of statistical mechanics, in the following way. In equilibrium, the free energy A of an assembly of N molecules is known to be given by

$$e^{-A/kT} = \frac{1}{N!}\int^{(2N)}\cdots\int e^{-W/kT}\prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)}, \tag{2·1}$$

where

$$W = \tfrac{1}{2}m\sum_{i=1}^{N}\boldsymbol{\xi}^{(i)2}+\Phi, \tag{2·2}$$

in the notation of part I. An independent proof of this result, based on the exact formula for n_2, will be found in the appendix. Note here that (2·2) implies that the molecules possess no internal degrees of freedom, a simplification which will be made throughout the present paper.

---

* The liquid phase is distinguished from the gaseous phase by the existence of real roots of a certain transcendental equation. In this respect the theory resembles a theory of condensation proposed by Kaplan & Dresden (1944). In spite of its attractive simplicity, however, the latter remains extremely speculative.

It was found in part I that, for equilibrium,

$$Qf_N = e^{-W/kT}, \tag{2·3}$$

where

$$Q = \frac{1}{N!}\int^{(2N)}\!\!\cdots\!\int e^{-W/kT}\prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)}, \tag{2·4}$$

so that

$$\begin{aligned}-\frac{A}{kT} &= \log Q \\ &= \frac{1}{N!}\int^{(2N)}\!\!\cdots\!\int f_N \log Q \prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)} \\ &= -\frac{1}{N!}\int^{(2N)}\!\!\cdots\!\int \left(\frac{W}{kT}+\log f_N\right) f_N \prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)}.\end{aligned} \tag{2·5}$$

The entropy is

$$S = \frac{E-A}{T}, \tag{2·6}$$

where

$$E = \frac{1}{N!}\int^{(2N)}\!\!\cdots\!\int W f_N \prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)} \tag{2·7}$$

is the internal energy. Hence, in equilibrium,

$$S = -\frac{k}{N!}\int^{(2N)}\!\!\cdots\!\int f_N \log f_N \prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)}. \tag{2·8}$$

(2·8) will be adopted as the definition of the entropy, even when equilibrium conditions do not exist, and the proof of the H-theorem will then consist of showing that

$$H = \int^{(2N)}\!\!\cdots\!\int f_N \log f_N \prod_{i=1}^{N} d\mathbf{x}^{(i)}\,d\boldsymbol{\xi}^{(i)} \tag{2·9}$$

cannot increase.

According to (4·5) of part I, the equation satisfied by f_N is

$$\frac{\partial f_N}{\partial t} + \sum_{i=1}^{N} \frac{\partial f_N}{\partial \mathbf{x}^{(i)}}\boldsymbol{\xi}^{(i)} = \frac{1}{m}\sum \frac{\partial f_N}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \Phi}{\partial \mathbf{x}^{(i)}}, \tag{2·10}$$

if f_{N+1} is assumed to vanish. The 'auxiliary equations' for the integration of (2·10),

$$\frac{dt}{1} = d\mathbf{x}^{(i)}/\boldsymbol{\xi}^{(i)} = -m\,d\boldsymbol{\xi}^{(i)}\Big/\frac{\partial \Phi}{\partial \mathbf{x}^{(i)}} \tag{2·11}$$

are identical with Lagrange's equations for the motion of the system. They yield $6N$ functionally independent integrals of the form

$$I_r(t, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}, \boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(N)}) = C_r, \quad r = 1, \ldots, 6N,$$

which, when solved for $\mathbf{x}^{(i)}$ and $\boldsymbol{\xi}^{(i)}$, provide a detailed description of the motion in terms of the time t and the parameters C_r which are specified by the initial conditions. f_N is thereby determined as an arbitrary function of the $6N$ variables I_r, which include the total energy W, the total momentum $\mathbf{M}$, and the total angular momentum $\mathbf{A}$ of the system. This solution is, however, much too general for the present purpose, which requires the form assumed by f_N when the system has been left for some time, and has arrived at a 'normal' state. Also, the idea of statistical interaction has disappeared from the equations, which, as one might expect, leads to the paradoxical result $\partial H/\partial t = 0$ even when statistical equilibrium has not been reached.

Therefore, instead of (2·10) the equation

$$\frac{\partial f_N}{\partial t}+\sum_{i=1}^{N}\frac{\partial f_N}{\partial \mathbf{x}^{(i)}}\boldsymbol{\xi}^{(i)}=\frac{1}{m}\sum_{i=1}^{N}\frac{\partial f_N}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \Phi}{\partial \mathbf{x}^{(i)}}+\frac{1}{m}\iint\sum_{i=1}^{N}\frac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \phi^{(iN+1)}}{\partial \mathbf{x}^{(i)}}\,d\mathbf{x}^{(N+1)}\,d\boldsymbol{\xi}^{(N+1)} \tag{2·12}$$

will be considered, which has the effect of providing an external system (to be imagined as the vapour with which the liquid is in contact, for example) with similar distribution properties, and with which the given system may statistically interact. Also, to make the treatment as general as possible, the possibility of a conservative field of force $-\partial\Psi/\partial\mathbf{x}$ acting on the molecules is contemplated, which requires only the replacement of Φ by $\Phi' = \Phi+\sum_{i=1}^{N}\Psi(\mathbf{x}^{(i)})$ in (2·12). This innovation has the advantage that the boundary of the fluid may be represented by a wall where $\Psi\to\infty$; of course, at the wall, f_N and surface integrals involving f_N must vanish, otherwise the total potential energy of the fluid would diverge.

3. The generalized H-theorem

Now proceed to calculate the rate of change with time of the quantity H defined in (2·9). Thus

$$\begin{aligned}\frac{\partial H}{\partial t}&=\int^{(2N)}\!\!\cdots\int(1+\log f_N)\frac{\partial f_N}{\partial t}\prod_{i=1}^{N}d\mathbf{x}^{(i)}d\boldsymbol{\xi}^{(i)}\\&=\frac{1}{m}\int^{(2N+2)}\!\!\cdots\int(1+\log f_N)\sum_{i=1}^{N}\frac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \phi^{(iN+1)}}{\partial \mathbf{x}^{(i)}}\prod_{j=1}^{N+1}d\mathbf{x}^{(j)}d\boldsymbol{\xi}^{(j)},\end{aligned} \tag{3·1}$$

since
$$\int^{(2N)}\!\!\cdots\int\boldsymbol{\xi}^{(i)}\frac{\partial}{\partial \mathbf{x}^{(i)}}(f_N\log f_N)\prod_{j=1}^{N}d\mathbf{x}^{(j)}d\boldsymbol{\xi}^{(j)}$$

and
$$\int^{(2N)}\!\!\cdots\int\frac{\partial \Phi}{\partial \mathbf{x}^{(i)}}\frac{\partial}{\partial \boldsymbol{\xi}^{(i)}}(f_N\log f_N)\prod_{j=1}^{N}d\mathbf{x}^{(j)}d\boldsymbol{\xi}^{(j)}$$

both vanish on transformation to surface integrals. f_N and f_{N+1} are symmetrical functions of their positions and velocities, so all terms under the summation in (3·1) contribute the same value, and hence

$$\frac{\partial H}{\partial t}=\frac{N}{m}\int^{(2N+2)}\!\!\cdots\int(1+\log f_N)\frac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(1)}}\frac{\partial \phi^{(1N+1)}}{\partial \mathbf{x}^{(1)}}\prod_{i=1}^{N+1}d\mathbf{x}^{(i)}d\boldsymbol{\xi}^{(i)}. \tag{3·2}$$

The analysis which follows is a generalization of that in appendix I to part I of this series, with which it should be compared. Write

$$\left.\begin{aligned}\mathbf{x} &= \tfrac{1}{2}(\mathbf{x}^{(1)}+\mathbf{x}^{(N+1)}), \quad \mathbf{r} = \mathbf{x}^{(N+1)}-\mathbf{x}^{(1)},\\ \boldsymbol{\xi} &= \tfrac{1}{2}(\boldsymbol{\xi}^{(1)}+\boldsymbol{\xi}^{(N+1)}), \quad \boldsymbol{\rho} = \boldsymbol{\xi}^{(N+1)}-\boldsymbol{\xi}^{(1)},\end{aligned}\right\} \tag{3·3}$$

$$\mathbf{m} = 2m\boldsymbol{\xi}, \quad \mathbf{a} = \tfrac{1}{2}m\mathbf{r}\times\boldsymbol{\rho}, \quad w = \tfrac{1}{4}m\rho^2+\phi(r). \tag{3·4}$$

By solving the equations (3·4) $\boldsymbol{\xi}$ and $\boldsymbol{\rho}$ can be expressed, and therefore $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(N+1)}$, in terms of $\mathbf{m}$, $\mathbf{a}$, w and $\mathbf{r}$*. A function $\bar{f}_{N+1}$ can then be defined by

$$f_{N+1} = \bar{f}_{N+1}(t, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N+1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N)}, \mathbf{m}, \mathbf{a}, w), \tag{3·5}$$

and it is easily verified that

$$\frac{1}{m}\left(\frac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(1)}}-\frac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(N+1)}}\right)\frac{\partial \phi}{\partial \mathbf{x}^{(1)}} = \frac{\partial \bar{f}_{N+1}}{\partial w}\boldsymbol{\rho}\frac{\partial \phi}{\partial \mathbf{r}} = \boldsymbol{\rho}\left(\frac{\partial f_{N+1}}{\partial \mathbf{x}^{(N+1)}}-\frac{\partial \bar{f}_{N+1}}{\partial \mathbf{x}^{(N+1)}}\right). \tag{3·6}$$

On substituting $\dfrac{1}{m}\dfrac{\partial f_{N+1}}{\partial \boldsymbol{\xi}^{(1)}}\dfrac{\partial \phi}{\partial \mathbf{x}^{(1)}}$ from (3·6) into (3·2) the only term which does not vanish is found to be

$$-N\int^{(2N+2)}\!\!\cdots\int(1+\log f_N)\,\boldsymbol{\rho}\,\frac{\partial \bar{f}_{N+1}}{\partial \mathbf{x}^{(N+1)}}\prod_{i=1}^{N+1}d\mathbf{x}^{(i)}d\boldsymbol{\xi}^{(i)}.$$

Now, $\mathbf{m}$, $\mathbf{a}$ and w may be regarded as geometrical parameters specifying the trajectories which would be followed by the molecules numbered (1) and $(N+1)$ if no other molecules were present. Thus, if the $\mathbf{x}^{(N+1)}$ domain be imagined as partitioned by tubes formed by the trajectories of $(N+1)$ relative to (1), it is seen that the integration over $\mathbf{x}^{(N+1)}$ may be performed, as in appendix I to part I, first along the typical trajectory, where $\mathbf{m}$, $\mathbf{a}$ and w are constant, and then over all values of the cross-section radius $\mathbf{b}$. At each end of the trajectory the interaction vanishes, and f_{N+1} factorizes into $f_1^{(N+1)}f_N$. Since the sphere around $\mathbf{x}^{(1)}$ in which ϕ is effectively different from zero is of microscopic dimensions, it need not be supposed that $\mathbf{x}^{(N+1)}$ differs macroscopically from either $\mathbf{x}^{(1)}$ or $\mathbf{x}$, or that the time differs from t, at the ends of the trajectories. The initial velocites $\boldsymbol{\xi}^{(1)\prime}$ and $\boldsymbol{\xi}^{(N+1)\prime}$ must, however, be determined from the actual, final velocities by the equations, following from (3·4)

$$\left.\begin{aligned}m(\boldsymbol{\xi}^{(1)\prime}+\boldsymbol{\xi}^{(N+1)\prime}) &= m(\boldsymbol{\xi}^{(1)}+\boldsymbol{\xi}^{(N+1)}),\\ m(\mathbf{x}^{(1)}\times\boldsymbol{\xi}^{(1)\prime}+\mathbf{x}^{(N+1)}\times\boldsymbol{\xi}^{(N+1)\prime}) &= m(\mathbf{x}^{(1)}\times\boldsymbol{\xi}^{(1)}+\mathbf{x}^{(N+1)}\times\boldsymbol{\xi}^{(N+1)}),\\ \tfrac{1}{2}m(\boldsymbol{\xi}^{(1)\prime 2}+\boldsymbol{\xi}^{(N+1)\prime 2}) &= \tfrac{1}{2}m(\boldsymbol{\xi}^{(1)2}+\boldsymbol{\xi}^{(N+1)2}).\end{aligned}\right\} \tag{3·7}$$

After the integration, (3·2) becomes

$$\frac{\partial H}{\partial t} = -N\int^{(2N+2)}\!\!\cdots\int(1+\log f_N)\,(f_N f_1^{(N+1)}-f_N' f_1^{(N+1)\prime})\,\rho\,d\mathbf{b}\,d\mathbf{x}\prod_{i=2}^{N}d\mathbf{x}^{(i)}\prod_{j=1}^{N+1}d\boldsymbol{\xi}^{(j)}. \tag{3·8}$$

---

* Only two components of **a** are strictly required for this transformation, but for symmetry this fact is ignored.

Now denote by F_N the function obtained from f_N by replacing the variables $\mathbf{x}^{(i)}$ and $\boldsymbol{\xi}^{(i)}$ by $\mathbf{x}^{(i+N)}$ and $\boldsymbol{\xi}^{(i+N)}$ respectively, and rewrite (3·8), using (2·4) of part I, in the form

$$\frac{\partial H}{\partial t} = -\frac{N}{(N-1)!}\int^{(4N)}\!\!\cdots\int (1+\log f_N)\,(f_N F_N - f'_N F'_N)\,\rho\, d\mathbf{b}\, d\mathbf{x} \prod_{i=2,\,N+2}^{N,\,2N} d\mathbf{x}^{(i)} \prod_{j=1}^{2N} d\boldsymbol{\xi}^{(j)}. \tag{3·9}$$

By a series of transformations familiar in the proof of the H-theorem for gases, one obtains finally

$$\frac{\partial H}{\partial t} = -\frac{N}{4(N-1)!}\int^{(4N)}\!\!\cdots\int \log\left\{\frac{f_N F_N}{f'_N F'_N}\right\}(f_N F_N - f'_N F'_N)\,\rho\, d\mathbf{b}\, d\mathbf{x} \prod_{i=2,\,N+2}^{N,\,2N} d\mathbf{x}^{(i)} \prod_{j=1}^{2N} d\boldsymbol{\xi}^{(j)}. \tag{3·10}$$

which makes it clear that $\partial H/\partial t$ is essentially negative or zero.

4. THE EXACT EQUILIBRIUM SOLUTION

It is inferred from the calculations of the last section that it is a condition of equilibrium that the entropy should have attained its maximum value, for which

$$\log f_N + \log F_N = \log f'_N + \log F'_N, \tag{4·1}$$

where

$$\left.\begin{aligned} F_N &= f_N(\mathbf{x}^{(N+1)}, \ldots, \mathbf{x}^{(2N)}, \boldsymbol{\xi}^{(N+1)}, \ldots, \boldsymbol{\xi}^{(2N)}),\\ f'_N &= f_N(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}, \boldsymbol{\xi}^{(1)\prime}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N)}),\\ F'_N &= F_N(\mathbf{x}^{(N+1)}, \ldots, \mathbf{x}^{(2N)}, \boldsymbol{\xi}^{(N+1)\prime}, \boldsymbol{\xi}^{(N+2)}, \ldots, \boldsymbol{\xi}^{(2N)}), \end{aligned}\right\} \tag{4·2}$$

and equations (3·7) are satisfied. Owing to the symmetry between the variables, one obtains from (4·1) by a succession of $N-1$ arbitrary transformations of the type (3·7),

$$\log f_N + \log F_N = \log f''_N + \log F''_N, \tag{4·3}$$

where

$$\begin{aligned} f''_N &= f_N(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}, \boldsymbol{\xi}^{(1)\prime}, \ldots, \boldsymbol{\xi}^{(N)\prime}),\\ F''_N &= f_N(\mathbf{x}^{(N+1)}, \ldots, \mathbf{x}^{(2N)}, \boldsymbol{\xi}^{(N+1)\prime}, \ldots, \boldsymbol{\xi}^{(2N)\prime}), \end{aligned} \tag{4·4}$$

and the only restrictions on the variables are

$$\left.\begin{aligned} \sum_{i=1}^{2N} \boldsymbol{\xi}^{(i)} &= \sum_{i=1}^{2N} \boldsymbol{\xi}^{(i)\prime},\\ \sum_{i=1}^{2N} \mathbf{x}^{(i)} \times \boldsymbol{\xi}^{(i)} &= \sum_{i=1}^{2N} \mathbf{x}^{(i)} \times \boldsymbol{\xi}^{(i)\prime},\\ \sum_{i=1}^{2N} \boldsymbol{\xi}^{(i)2} &= \sum_{i=1}^{2N} \boldsymbol{\xi}^{(i)\prime 2}. \end{aligned}\right\} \tag{4·5}$$

It follows that

$$\log f_N = C + m\mathbf{c}\sum_{i=1}^{N} \boldsymbol{\xi}^{(i)} + m\mathbf{c}'\sum_{i=1}^{N} \mathbf{x}^{(i)} \times \boldsymbol{\xi}^{(i)} + \tfrac{1}{2}mC'\sum_{i=1}^{N} \boldsymbol{\xi}^{(i)2}, \tag{4·6}$$

where C, $\mathbf{c}$, $\mathbf{c}'$ and C' may be functions of the $\mathbf{x}^{(i)}$.

To determine these functions, substitute in (2·12), modified for equilibrium conditions. Then

$$\frac{\partial f_N}{\partial t}+\sum_{i=1}^{N}\overline{\boldsymbol{\xi}^{(i)}}\frac{\partial f_N}{\partial \mathbf{x}^{(i)}}$$

vanishes; the statistical term

$$\frac{1}{m}\iint\sum_{i=1}^{N}\frac{\partial f_N}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \phi^{(iN+1)}}{\partial \mathbf{x}^{(i)}}\,d\mathbf{x}^{(N+1)}\,d\boldsymbol{\xi}^{(N+1)}$$

can be reduced to terms of the form

$$-\frac{1}{(N-1)!}\int^{(2N)}\cdots\int(f_N F_N-f'_N F'_N)\,\rho\,d\mathbf{b}\prod_{i=N+2}^{2N}d\mathbf{x}^{(i)}\prod_{j=N+1}^{2N}d\boldsymbol{\xi}^{(j)}$$

by analysis similar to that of the last section, and also vanishes. (2·12) therefore reduces to

$$\sum_{i=1}^{N}\frac{\partial f_N}{\partial \mathbf{x}^{(i)}}(\boldsymbol{\xi}^{(i)}-\overline{\boldsymbol{\xi}^{(i)}})=\frac{1}{m}\sum_{i=1}^{N}\frac{\partial f_N}{\partial \boldsymbol{\xi}^{(i)}}\frac{\partial \Phi}{\partial \mathbf{x}^{(i)}},\tag{4·7}$$

and then, on substitution from (4·6),

$$\left.\begin{aligned}\frac{\partial C'}{\partial \mathbf{x}^{(i)}}&=0,\\ \frac{\partial}{\partial \mathbf{x}^{(i)}}(\mathbf{c}+\mathbf{c}'\times\mathbf{x}^{(j)})+\frac{\partial}{\partial \mathbf{x}^{(j)}}(\mathbf{c}+\mathbf{c}'\times\mathbf{x}^{(i)})&=0,\\ \frac{\partial C}{\partial \mathbf{x}^{(i)}}&=C'\frac{\partial \Phi'}{\partial \mathbf{x}^{(i)}}+m\sum_{j=1}^{N}\overline{\boldsymbol{\xi}^{(j)}}\cdot\frac{\partial}{\partial \mathbf{x}^{(j)}}(\mathbf{c}+\mathbf{c}'\times\mathbf{x}^{(i)}),\\ \sum_{i=1}^{N}(\mathbf{c}+\mathbf{c}'\times\mathbf{x}^{(i)})\frac{\partial \Phi'}{\partial \mathbf{x}^{(i)}}+\sum_{i=1}^{N}\overline{\boldsymbol{\xi}^{(i)}}\frac{\partial C}{\partial \mathbf{x}^{(i)}}&=0.\end{aligned}\right\}\tag{4·8}$$

The first and second of these equations show that C', $\mathbf{c}$ and $\mathbf{c}'$ are all constants; C' is found to be given by $-1/kT$, when the temperature T is defined by

$$\tfrac{3}{2}kT=\frac{1}{2}\frac{m}{N!}\int^{(2N)}\cdots\int f_N\sum_{i=1}^{N}(\boldsymbol{\xi}^{(i)}-\overline{\boldsymbol{\xi}^{(i)}})^2\prod_{j=1}^{N}d\mathbf{x}^{(j)}d\boldsymbol{\xi}^{(j)}.\tag{4·9}$$

Also, by calculation, it is found that $\overline{\boldsymbol{\xi}^{(i)}}=kT(\mathbf{c}+\mathbf{c}'\times\mathbf{x}^{(i)})$, so that $\mathbf{c}$ and $\mathbf{c}'$ find simple interpretations as multiples of the velocity and angular velocity respectively; the fluid, if in motion, must behave macroscopically like a rigid body. The third of equations (4·8), taking account of the second, requires that

$$C=C'\left(\Phi'+\tfrac{1}{2}m\sum_{i=1}^{N}\overline{\boldsymbol{\xi}^{(i)2}}\right)+C'',$$

where C'' is a pure constant; and the last reduces to $\sum_{i,j=1}^{N}\overline{\boldsymbol{\xi}^{(i)}}\cdot\frac{\partial}{\partial \mathbf{x}^{(i)}}\overline{\boldsymbol{\xi}^{(j)2}}=0$, showing only that the visible kinetic energy $\tfrac{1}{2}m\sum_{j=1}^{N}\overline{\boldsymbol{\xi}^{(j)2}}$ is conserved.

The most general form for f_N when there are no changes in the entropy has thus been obtained.

5. THE SOLUTION OF THE INTEGRAL EQUATION

Now proceed to examine the equilibrium state in detail by the method proposed in § 5 of part I, namely, by the solution of the integral equation for $n_2(r)$:

$$\log\frac{n_2(r)}{n_1^2}+\frac{\phi(r)}{kT}=\frac{\pi}{n_1^3}\int_0^\infty\int_{-s}^{s}(s^2-t^2)\frac{t+r}{r}\{n_2(t+r)-n_1^2\}\,dt\,n_2(s)\frac{\phi'(s)}{kT}\,ds. \tag{5·1}$$

First linearize the equation by writing

$$n_2(r)=n_1^2e^{-\phi(r)/kT}\{1+f(r)\}, \tag{5·2}$$

and neglecting squares and higher powers of $f(r)$. This procedure will be justified if the right-hand side of (5·1) is not too large; hence

$$\begin{aligned}rf(r)=&-\pi n_1\int_0^\infty\int_{-s}^{s}(s^2-t^2)(t+r)f(t+r)\{1+\alpha(t+r)\}\,dt\,\alpha'(s)\{1+f(s)\}\,ds\\&-\pi n_1\int_0^\infty\int_{-s}^{s}(s^2-t^2)(t+r)\,\alpha(t+r)\,dt\,\alpha'(s)\{1+f(s)\}\,ds,\end{aligned} \tag{5·3}$$

where

$$\alpha(r)=e^{-\phi(r)/kT}-1. \tag{5·4}$$

Now observe that $\alpha(r)$ and $\alpha'(r)$ are both negligible except for small r, so that, when multiplied by these factors, $f(r)$ may be replaced by $\epsilon-1$, its average value in the neighbourhood of the origin. Then, after integration by parts,

$$rf(r)=2\pi n_1\int_0^\infty\int_{-s}^{s}(t+r)\{f(t+r)+\epsilon\alpha(t+r)\}\,dt\,\epsilon\alpha(s)\,s\,ds. \tag{5·5}$$

This equation can be solved by Fourier transforms; writing

$$\left.\begin{aligned}rg(r)&=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}sf(s)\sin(rs)\,ds,\\ r\beta(r)&=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}s\alpha(s)\sin(rs)\,ds,\end{aligned}\right\} \tag{5·6}$$

then (5·5) becomes

$$\lambda g(r)=\{g(r)+\epsilon\beta(r)\}\,\epsilon\beta(r), \tag{5·7}$$

where $\lambda^{-1}=(2\pi)^{\frac{1}{2}}n_1$, so that

$$rf(r)=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{\epsilon^2s\beta(s)^2\sin(rs)\,ds}{\lambda-\epsilon\beta(s)}. \tag{5·8}$$

It will be supposed that there is an analytic function $\beta(z)$ which reduces to $\beta(r)$ when $z=r$ is real. The most general solution of (5·5) is then given by (5·8), where the path of integration may follow any course from negative to positive infinity in the complex plane. This is most easily seen by remarking that if $f(r)$ is a solution, so is $f(r)+\frac{A_k}{r}\sin(rz_k)$, where z_k is any root of the transcendental equation

$$\beta(z)=\lambda/\epsilon; \tag{5·9}$$

the additional term may, however, be obtained from (5·8) by deforming the path of integration to enclose the singular point z_k. It will be seen in the next section, however, that the precise path of integration is determined by the boundary condition

$$f(r) = O(r^{-3}) \quad \text{as} \quad r \to \infty, \tag{5·10}$$

which follows from the requirement that $\int (n_2^{(12)} - n_1^2)\, dx^{(2)}$ should converge. (Strictly, one should write $f(r) = O(r^{-(3+\delta)})$, $\delta > 0$, but owing to the approximate nature of (5·2), this refinement is not compulsory.)

6. THERMODYNAMICAL EXPRESSIONS

It is now possible to calculate explicitly the internal energy E, given by

$$E = \tfrac{1}{2}V \int_0^\infty n_2(r)\,\phi(r)\,4\pi r^2 dr + \tfrac{3}{2}NkT, \tag{6·1}$$

and the pressure p from

$$p = n_1 kT - \tfrac{1}{6}\int_0^\infty n_2(r)\,\phi'(r) r 4\pi r^2 dr. \tag{6·2}$$

These formulae are known, but a comprehensive derivation is included in the appendix. Then, from (5·2) and (5·8),

$$E = \frac{NkT^2}{2\lambda}\left[\frac{\partial}{\partial T}\{\beta(0)\} + \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{\epsilon^2\beta(s)^2\dfrac{\partial\{\beta(s)\}}{\partial T}s^2 ds}{\lambda - \epsilon\beta(s)}\right] + \tfrac{3}{2}NkT, \tag{6·3}$$

and

$$p = n_1 kT\left\{1 + \frac{n_1 I}{6(2\pi)^{\frac{1}{2}}\lambda}\right\}, \tag{6·4}$$

where

$$\begin{aligned} I &= \int_{-\infty}^{\infty}\alpha'(r)\{1+f(r)\}r^3 dr \\ &= -\int_{-\infty}^{\infty}\alpha(r)\,[3\{1+f(r)\}+rf'(r)]\,r^2 dr \\ &= -3(2\pi)^{\frac{1}{2}}\beta(0) - \int_{-\infty}^{\infty}\frac{\epsilon^2\beta(s)^2\{3\beta(s)+s\beta'(s)\}s^2 ds}{\lambda-\epsilon\beta(s)} \\ &= -3(2\pi)^{\frac{1}{2}}\beta(0) - \int_{-\infty}^{\infty}\left[-\tfrac{1}{2}\epsilon\beta(s)^2 + \frac{\lambda^2\beta(s)}{\lambda-\epsilon\beta(s)} + \frac{\lambda^2}{\epsilon}\log\left|1-\frac{\epsilon}{\lambda}\beta(s)\right|\right]s^2 ds. \end{aligned} \tag{6·5}$$

Combining (6·3) and (6·4), it is confirmed that the equation

$$d\left(\frac{A}{T}\right) = -\frac{E}{T^2}dT - \frac{pdV}{T} \tag{6·6}$$

is integrable, yielding for the free energy

$$\begin{aligned} A = \frac{NkT}{2\lambda}\left[-\beta(0) + \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\left[\tfrac{1}{2}\epsilon\beta(s)^2 + \lambda\beta(s) + \frac{\lambda^2}{\epsilon}\log\left|1-\frac{\epsilon}{\lambda}\beta(s)\right|\right]s^2 ds\right. \\ \left. - \tfrac{3}{2}NkT\log T + NkT\log n_1 \right. \end{aligned} \tag{6·7}$$

apart from an insignificant multiple of T.

Explicit formulae for all the important thermodynamical functions have thus been obtained, except that the path of integration, which is the same in (5·8), (6·3), (6·5) and (6·7), is so far undetermined. To apply the boundary condition use is made of the well-known lemma

$$\lim_{r\to\infty}\int_{-\infty}^{\infty}\frac{\psi(s)}{s}\sin(rs)\,ds = \lim_{t\to 0}\pi\psi(t). \tag{6·8}$$

It is necessary also to discuss the dependence of $\alpha(r)$ and $\beta(r)$ on the temperature. Since $\phi(r)$ is negative for all large r, in this range $\alpha(r)$ will become increasingly positive as the temperature decreases. It follows that $\beta(r)$ will behave in the same way when r is small. Referring to (5·9), it is necessary to distinguish between two cases. For sufficiently low densities and high temperatures, λ/ϵ exceeds the upper bound of $\beta(r)$, and (5·9) has no real roots; the path of integration may then be taken along the real axis, since (5·10) is thereby satisfied. In the region of higher densities and lower temperatures, however, (5·9) will have a pair of equal and opposite real roots (since $\beta(z)$ is an even function of z), and in this case the path of integration must be taken just above both singularities on the real axis. On such a path,

$$\int_{-\infty}^{\infty}\frac{s\psi(s)\sin(rs)\,ds}{s^2-s_0^2} = \int_{-\infty}^{\infty}\frac{s\{\psi(s)-\psi(s_0)\}\sin(rs)\,ds}{s^2-s_0^2},$$

so that the singularities in the integrand do not affect the behaviour of the integral for large r.

The possibility that *the two cases described above correspond to the gaseous and liquid phases respectively* has already been expressed, and in the following sections convincing evidence will be given to confirm this view.

7. DEVELOPMENT FOR THE GASEOUS PHASE

The expressions found in the last section for the thermodynamical functions are quite general, but not very informative to the physical intuition, or convenient for numerical evaluation. Certain developments in series of the general expressions will therefore be considered. In the region of low densities and high temperatures $\left|\frac{\epsilon\beta(s)}{\lambda}\right| < 1$ for all real s, and it is permissible to develop the infinite integrals in powers of the density. Hence

$$A = -NkT\left\{\frac{\beta(0)}{2\lambda}+\sum_{\nu=2}^{\infty}\frac{\epsilon^\nu}{2(2\pi)^{\frac{1}{2}}\lambda^\nu}\int_{-\infty}^{\infty}\frac{\beta(s)^{\nu+1}s^2ds}{\nu+1}\right\}$$
$$-\tfrac{3}{2}NkT\log T+NkT\log n_1 \tag{7·1}$$

and

$$p = n_1kT\left\{1-\frac{\beta(0)}{2\lambda}-\sum_{\nu=2}^{\infty}\frac{\nu\epsilon^\nu}{2(2\pi)^{\frac{1}{2}}\lambda^\nu}\int_{-\infty}^{\infty}\frac{\beta(s)^{\nu+1}s^2ds}{\nu+1}\right\}, \tag{7·2}$$

which is the equation of state.

Under the same conditions, however, Mayer (1937) has obtained an exact development quite analogous to (7·1) by evaluation of the integral expression in (2·1). A concise and perfected exposition of the method has been given by Born & Fuchs (1938); the result obtained is

$$A = -NkT\sum_{\nu=1}^{\infty}\frac{\beta_\nu n_1^\nu}{\nu+1} - \tfrac{3}{2}NkT\log T + NkT\log n_1, \tag{7·3}$$

where the first β_ν are given by

$$\left.\begin{aligned} \beta_1 &= \int_0^\infty \alpha(r)\,4\pi r^2\,dr, \\ \beta_2 &= \tfrac{1}{2}\iint \alpha^{(12)}\alpha^{(23)}\alpha^{(31)}\,d\mathbf{x}^{(2)}\,d\mathbf{x}^{(3)}, \\ \alpha^{(rs)} &= \alpha(|\mathbf{x}^{(s)}-\mathbf{x}^{(r)}|), \end{aligned}\right\} \tag{7·4}$$

or

$$\left.\begin{aligned} \beta_1 &= (2\pi)^{\frac{3}{2}}\beta(0), \\ \beta_2 &= \tfrac{1}{2}(2\pi)^{\frac{5}{2}}\int_0^\infty \beta(y)^3\,4\pi y^3\,dy, \\ \alpha^{(rs)} &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int e^{i\mathbf{y}\,(\mathbf{x}^{(s)}-\mathbf{x}^{(r)})}\,\beta(y)\,d\mathbf{y}. \end{aligned}\right\} \tag{7·5}$$

Comparing (7·1) and (7·3), it is found that

$$\beta_\nu = \tfrac{1}{2}\epsilon^\nu(2\pi)^{\frac{1}{2}(3\nu-1)}\int_{-\infty}^{\infty}\beta(s)^{\nu+1}\,s^2\,ds. \tag{7·6}$$

Substituting $\nu = 2$ in (7·6), agreement is obtained with the value of β_2 given by (7·5), provided $\epsilon = 1$; the remaining β_ν given by (7·6) are approximate, but far more tractable expressions for the virial coefficients than those obtained by Mayer's combinational method.

Mayer found that the equation of state derived from (7·3) will not describe the liquid phase, for the simple reason that the infinite series becomes divergent on approaching the point of condensation. Precisely the same thing happens to the present series (7·1) when the transcendental equation (5·9) first has a real root, and the provisional hypothesis advanced at the end of the last section is therefore confirmed.

It is important to notice, however, that the solution of the equation (5·9) will not necessarily yield the boiling-point as a function of density, but rather the highest temperature to which the liquid can be superheated without boiling. The boiling-point must be determined by the usual thermodynamical method of considering the equilibrium between the two phases.*

---

* The authors of this series believe that the view of Kahn (1938) on this matter is incorrect, and will give detailed reasons later.

8. Development for the liquid phase

It has been seen that the power series (7·1) can describe only the gaseous phase, and that another form of expansion is required to adapt the formulae of § 6 for the liquid as well as the gas. For this purpose it is convenient to consider the complex roots z_u ($u = 1, 2, 3,$ etc.) of (5·9) in the upper half of the z-plane. Assume that $\beta(z) \to 0$ as $|z| \to \infty$, uniformly with respect to $\arg z$, in this region; by the theory of residues, the following expansion is then obtained from (5·8):

$$r\{f(r) + \epsilon\alpha(r)\} = -(2\pi)^{\frac{1}{2}}\frac{\lambda^2}{\epsilon}\sum_u \frac{z_u e^{ivz_u}}{\beta'(z_u)}, \tag{8·1}$$

where, according to (5·6),

$$z_u\beta'(z_u) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} \cos(rz_u)\,\alpha(s)\,s^2 ds - \lambda/\epsilon, \tag{8·2}$$

so that
$$-z_u\beta'(z_u)\,kT^2\frac{\partial z_u}{\partial T} = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} \sin(rz_u)\,e^{-\phi(s)/kT}\,\phi(s)\,s\,ds. \tag{8·3}$$

Now use (8·1) to calculate E afresh from (6·1), and obtain

$$E = \frac{NkT^2}{2\lambda}\left[\frac{\partial}{\partial T}\left\{\beta(0) - \frac{\epsilon}{2(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\beta(s)^2 s^2 ds\right\} + (2\pi)^{\frac{1}{2}}\,i\frac{\lambda^2}{\epsilon}\sum_u z_u^2\frac{\partial z_u}{\partial T}\right] + \tfrac{3}{2}NkT. \tag{8·4}$$

The equation of state, obtained similarly from (6·2), is

$$p = n_1 kT\left[1 - \frac{\beta(0)}{2\lambda} + \frac{\epsilon}{4(2\pi)^{\frac{1}{2}}\lambda}\int_{-\infty}^{\infty}\beta(r)^2 r^2 dr + \frac{(2\pi)^{\frac{1}{2}}i\lambda}{2\epsilon}\sum_u\left(\frac{\lambda z_u^2}{\epsilon\beta'(z_u)} + \frac{z_u^3}{3}\right)\right]. \tag{8·5}$$

Since
$$\beta'(z_u)\frac{\partial z_u}{\partial V} = \frac{\lambda n_1}{\epsilon N}, \tag{8·6}$$

the equation (6·6) is again integrable, and

$$A = -\frac{NkT}{2\lambda}\left\{\beta(0) - \frac{\epsilon}{2(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\beta(r)^2 r^2 dr + \frac{(2\pi)^{\frac{1}{2}}i\lambda^2}{3\epsilon}\sum_u z_u^3\right\} - \tfrac{3}{2}NkT\log T + NkT\log n_1. \tag{8·7}$$

The formulae of this section hold equally for the liquid and the gas. The difference arises from the fact that in the case of the liquid there are two real roots which are not included in the u-summation; when the temperature is raised beyond the boiling-point these roots migrate in opposite directions into the complex plane, and the one above the real axis is now included in the u-summation. Thus there is a real discontinuity in the radial distribution function and all the thermodynamic functions at the point of condensation—a consequence of the theory which is confirmed by experience. However, the discontinuity in the radial distribution function corresponds only to a transfer from one branch of the general solution of the integral equation to another; the point of condensation is therefore a kind of branch point.

An approximate classical theory of the solid state appears to result in the following way. When the boundary condition (5·10), which can be correct only for liquids, is omitted, the possibility of a periodic term $\frac{Ak}{r}\sin rz_k$ in the expression for $n_2(r)$, where z_k is the real root of (5·9), is not excluded. Such periodicity at large distances from a given lattice point is characteristic of the solid state. This may provide confirmation of the idea of Vlasov mentioned in the introduction to the first paper of this series.

9. APPENDIX

Here a proof of the formula (2·1) for the free energy will be given. First show that the quantity there defined satisfies the differential equation (6·6), or the equivalent partial differential equations

$$\frac{\partial}{\partial T}\left(\frac{A}{T}\right) = -\frac{E}{T^2}; \quad \frac{\partial}{\partial V}\left(\frac{A}{T}\right) = -\frac{p}{T}, \tag{9·1}$$

where E and p are defined by (6·1) and (6·2) respectively. Use is made of the formula

$$Q_1 n_2 = \frac{1}{(N-2)!}\int^{(N-2)}\cdots\int e^{-\Phi/kT}\prod_{i=3}^{N} d\mathbf{x}^{(i)}, \tag{9·2}$$

where

$$\begin{aligned} Q_1 &= \frac{1}{N!}\int^{(N)}\cdots\int e^{-\Phi/kT}\prod_{i=1}^{N} d\mathbf{x}^{(i)} \\ &= \left(\frac{2\pi kT}{m}\right)^{-\frac{1}{2}(3N)} e^{-A/kT}. \end{aligned} \tag{9·3}$$

The first of the equations (9·1) follows immediately on differentiating (9·3); to obtain the second, write $\mathbf{x}^{(i)} = l\boldsymbol{\theta}^{(i)}$, where $l^3 = V$; then

$$\left.\begin{aligned} Q_1 &= \frac{l^{3N}}{N!}\int_{-\frac{1}{2}}^{\frac{1}{2}(3N)}\cdots\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-\Phi/kT}\prod_{i=1}^{N} d\theta_1^{(i)}\,d\theta_2^{(i)}\,d\theta_3^{(i)}, \\ \Phi &= \tfrac{1}{2}\sum_{i,j=1}^{N}\phi\{l\,|\boldsymbol{\theta}^{(j)}-\boldsymbol{\theta}^{(i)}|\}, \end{aligned}\right\} \tag{9·4}$$

so that

$$\frac{\partial Q_1}{\partial l} = \frac{3NQ_1}{l} - \frac{l^{3N}}{2(N-2)!}\int_{-\frac{1}{2}}^{\frac{1}{2}(3N)}\cdots\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-\Phi/kT}\frac{\phi'\{l\,|\boldsymbol{\theta}^{(2)}-\boldsymbol{\theta}^{(1)}|\}}{kT}\,|\boldsymbol{\theta}^{(2)}-\boldsymbol{\theta}^{(1)}|\prod_{i=1}^{N} d\theta_1^{(i)}\,d\theta_2^{(i)}\,d\theta_3^{(i)} \tag{9·5}$$

and

$$\begin{aligned} \frac{\partial Q_1}{\partial V} &= \frac{1}{3l^2}\frac{\partial Q_1}{\partial l} \\ &= n_1 Q_1 - \tfrac{1}{6} Q_1 \int n_2(r)\frac{\phi'(r)}{kT}\,r\,d\mathbf{r}. \end{aligned} \tag{9·6}$$

This leads immediately to the second of equations (9·1).

It remains to be shown that the quantities E and p as defined correspond to the internal energy and pressure as observed experimentally. That this is so in the case of the internal energy follows easily on calculating the mean value of the Hamiltonian energy W. The evaluation of the pressure requires rather more attention; in order to make the discussion as general as possible, equilibrium conditions will not be assumed, and the pressure tensor in the fluid will be evaluated in the first instance. (The analysis which follows is borrowed from part III on dynamical properties, now in preparation by Professor M. Born and the author.)

Multiply the equation

$$\frac{\partial f_1}{\partial t}+\boldsymbol{\xi}^{(1)}\frac{\partial f_1}{\partial \mathbf{x}^{(1)}}=\frac{1}{m}\iint\frac{\partial \phi}{\partial \mathbf{x}^{(1)}}\frac{\partial f_2}{\partial \boldsymbol{\xi}^{(1)}}\,d\mathbf{x}^{(2)}\,d\boldsymbol{\xi}^{(2)}, \tag{9·7}$$

which is satisfied by f_1 in the absence of external forces, by $m\boldsymbol{\xi}^{(1)}$, and integrate over $\boldsymbol{\xi}^{(1)}$, obtaining

$$\begin{aligned}\frac{\partial}{\partial t}\left\{\int m\xi^{(1)}_{\alpha}f_1\,d\boldsymbol{\xi}^{(1)}\right\}+\sum_{\beta=1}^{3}\frac{\partial}{\partial x^{(1)}_{\beta}}\left\{\int m\xi^{(1)}_{\alpha}\xi^{(1)}_{\beta}f_1\,d\boldsymbol{\xi}^{(1)}\right\}&\\ =\sum_{\beta=1}^{3}\iiint\frac{\partial \phi}{\partial x^{(1)}_{\beta}}\frac{\partial f_2}{\partial \xi^{(1)}_{\beta}}\xi^{(1)}_{\alpha}\,d\mathbf{x}^{(2)}\,d\boldsymbol{\xi}^{(2)}\,d\boldsymbol{\xi}^{(1)}&\\ =-\int\frac{\partial \phi}{\partial x^{(1)}_{\alpha}}n_2\,d\mathbf{x}^{(2)}.&\end{aligned} \tag{9·8}$$

To simplify further the right-hand side, note that, although macroscopically uniform conditions are not yet assumed, within the microscopic sphere around $\mathbf{x}^{(2)}$, where $\partial\phi/\partial\mathbf{x}^{(1)}$ is effectively different from zero, n_2 may be represented as a function of $r=|\mathbf{x}^{(2)}-\mathbf{x}^{(1)}|$, and writing

$$\nu(\mathbf{x}^{(2)},r)=\int_{\infty}^{r}n_2(\mathbf{x}^{(2)},r)\,\phi'(r)\,dr, \tag{9·9}$$

then (9·8) becomes

$$\frac{\partial}{\partial t}(mn_1\xi^{(1)}_{\alpha})+\sum_{\beta=1}^{3}\frac{\partial}{\partial x^{(1)}_{\beta}}\left\{\int m\xi^{(1)}_{\alpha}\xi^{(1)}_{\beta}f_1\,d\boldsymbol{\xi}^{(1)}+\int\nu\,d\mathbf{x}^{(2)}\delta_{\alpha\beta}\right\}=0. \tag{9·10}$$

Further, since

$$\frac{\partial}{\partial t}(n_1\xi^{(1)}_{\alpha})=n_1\frac{\overline{d\xi^{(1)}_{\alpha}}}{dt}-\sum_{\beta=1}^{3}\frac{\partial}{\partial x^{(1)}_{\beta}}(n_1\overline{\xi^{(1)}_{\alpha}}\,\overline{\xi^{(1)}_{\beta}}), \tag{9·11}$$

where d/dt is the operator $\dfrac{\partial}{\partial t}+\sum_{\beta=1}^{3}\overline{\xi^{(1)}_{\beta}}\dfrac{\partial}{\partial x^{(1)}_{\beta}}$, (9·10) reduces to

$$mn_1\frac{\overline{d\xi^{(1)}_{\alpha}}}{dt}+\sum_{\beta=1}^{3}\frac{\partial p_{\alpha\beta}}{\partial x^{(1)}_{\beta}}=0, \tag{9·12}$$

where

$$p_{\alpha\beta}=m\int(\xi^{(1)}_{\alpha}\xi^{(1)}_{\beta}-\overline{\xi^{(1)}_{\alpha}}\,\overline{\xi^{(1)}_{\beta}})f_1\,d\boldsymbol{\xi}^{(1)}+\int\nu\,d\mathbf{x}^{(2)}\delta_{\alpha\beta}+\pi_{\alpha\beta}, \tag{9·13}$$

and $\pi_{\alpha\beta}$ is any tensor which satisfies $\sum_{\beta=1}^{3} \frac{\partial \pi_{\alpha\beta}}{\partial x_\beta^{(1)}} = 0$. $\pi_{\alpha\beta}$ must be determined by conditions at the boundary of the fluid, where the part of $p_{\alpha\beta}$ due to the molecular forces, namely,

$$\int \nu(\mathbf{x}^{(1)} + \mathbf{r}, r)\, d\mathbf{r}\, \delta_{\alpha\beta} + \pi_{\alpha\beta},$$

must vanish. (The interaction between the walls and the fluid is neglected.)

In uniform conditions, $\nu(\mathbf{x}^{(2)}, r)$ depends only on r, and $\pi_{\alpha\beta}$ is a constant tensor, given by

$$\pi_{\alpha\beta} = -\tfrac{1}{2}\int_0^\infty \nu(r)\, 4\pi r^2 dr\, \delta_{\alpha\beta}, \tag{9·14}$$

since $\mathbf{r}$ ranges only over a hemisphere at the boundary. The hydrostatic pressure in equilibrium conditions is therefore given by

$$\begin{aligned} p &= \tfrac{1}{3}\int \xi^{(1)2} f_1\, d\xi^{(1)} + \tfrac{1}{2}\int_0^\infty \nu(r)\, 4\pi r^2 dr \\ &= n_1 kT - \tfrac{1}{6}\int_0^\infty n_2(r)\, \phi'(r)\, r\, 4\pi r^2 dr, \end{aligned} \tag{9·15}$$

in agreement with the value found previously by differentiating the phase integral.

The author wishes to express here his grateful appreciation of numerous helpful discussions with Professor Born in the course of this work. He is also indebted to Dr R. Fürth for some useful comments.

REFERENCES

Born, M. & Fuchs, K. 1938 *Proc. Roy. Soc.* A, **166**, 391.
Born, M. & Green, H. S. 1946 *Proc. Roy. Soc.* A, **188**, 10.
Chapman, S. & Cowling, T. G. 1939 *The mathematical theory of non-uniform gases*. Camb. Univ. Press.
Kahn, B. 1938 *Utrecht Diss.*
Kaplan, W. & Dresden, M. 1944 *Phys. Rev.* **66**, 16.
Mayer, J. E. 1937 *J. Chem. Phys.* **5**, 67, 74; **6**, 87, 101.

Tables of second virial and low-pressure Joule-Thomson coefficients for intermolecular potentials with exponential repulsion

By R. A. Buckingham, Ph.D., *University College, London*

and J. Corner, Ph.D., *Armament Research Department, Ministry of Supply*

(*Communicated by H. S. W. Massey, F.R.S.—Received* 6 *June* 1946)

The potentials used are functions of r, the distance between the centres of the molecules, but not of their relative orientation. The long-range attraction between the molecules is taken to be the sum of terms proportional to r^{-6} and r^{-8}, and the short-range repulsion is proportional to $e^{-\rho r}$. This potential has four parameters: the position and depth of the minimum of the potential, the relative importance of the r^{-8} term, and the steepness of the exponential repulsion. The second virial and low-pressure Joule-Thomson coefficients are tabulated for the range of parameters likely to be found in actual molecules, and for temperatures from the critical temperature to about 20 times this value. The most important of the three quantum corrections is tabulated over the same ranges.

INTRODUCTION

The second virial of a gas is connected with the field of force round its molecules by an integral which is simple in form but in general difficult to evaluate analytically. This has led to a search for types of intermolecular potential energy which would have the right qualitative features and would give an analytical formula for the second virial coefficient. The most successful of these forms has been the Lennard-Jones potential

$$E(r) = \lambda r^{-n} - \mu r^{-m}, \qquad (1)$$

where r is the distance between the centres of the molecules, λ, μ, n and m are constants, and $n > m > 3$. This has been integrated as a convergent series (Lennard-Jones 1924), which is convenient for the analysis of observed second virial coefficients. This potential has been applied to many substances.

More recently, several authors have found that in particular cases the repulsive part of (1), namely, λr^{-n}, can with advantage be replaced by an exponential repulsion $Ae^{-\rho r}$ (Born & Mayer 1932; Bleick & Mayer 1934; Huggins 1937). Unfortunately, the second virial coefficient of the resulting potential has not yet been evaluated analytically, and it has been necessary to use numerical integration (cf. Herzfeld 1937) or approximate methods (Buckingham 1938). Recently, Orr (1942) has proposed an intermolecular potential of rather different form which has the desired feature of an exponential repulsion at short distances, and whose virial coefficient can be integrated exactly.

We have computed the second virial coefficient for a potential derived from the usual type (1) by replacing the inverse-power repulsion λr^{-n} by an exponential term $Ae^{-\rho r}$, taking $m = 6$, and adding an additional attractive term proportional to r^{-8}. Reduced to non-dimensional variables, of which there are three, the results previously obtained for simple molecules show that one need compute the second virial coefficient for only quite restricted ranges of two of the variables. This makes the computation not intolerably long, though still rather laborious. The tables of second virial coefficients are given in the present paper; their application to various substances will follow in a separate paper.

Hirschfelder, Ewell & Roebuck (1938) have shown how the Joule-Thomson coefficient (extrapolated to zero pressure) may be used instead of the second virial coefficient to find intermolecular potentials, and they have derived the Joule-Thomson coefficient of the potential (1) from Lennard-Jones's expression for its second virial coefficient. We have prepared tables of the Joule-Thomson coefficient for the same potentials for which we have computed the second virial coefficient. We have also calculated similar tables of the most important of the three quantum corrections to the second virial and Joule-Thomson coefficients.

The potentials used in this paper can apply only to molecules which are nearly spherical.

FORMULAE USED

The second virial coefficient has been defined in several slightly different ways, and has been expressed in several sets of units (cf. Fowler 1936, and Hirschfelder *et al.* 1938). We use the definition

$$pV = nkT\left\{1+\frac{B}{V}+0(V^{-2})\right\}, \tag{2}$$

where p, T and n are the pressure, temperature and number of molecules in the volume V. (2) is true at low densities, and defines the second virial coefficient B, which is a function of temperature. B is proportional to n, and has the dimensions of a volume. The most usual choices of n are $n = N$, Avogadro's number, so that B is given as volume per gram-molecule, or $n = L$, Loschmidt's number, which means that B is in units of volume (per c.c. of perfect gas at S.T.P.). n need not be specified in this paper.

Take the zero of the intermolecular potential energy $E(r)$ to be the value for infinite r. In classical statistical mechanics, B is connected with $E(r)$ by the relation

$$B = 2\pi n\int_0^\infty \{1-e^{-E(r)/kT}\}r^2\,dr. \tag{3}$$

There are three quantum corrections:

(I) the probability of configurations of potential energy E is not proportional to $e^{-E/kT}$;

(II) certain states of the gas are excluded by the statistics appropriate to the molecules under consideration;

(III) when the kinetic energy of the relative motion of two molecules is less than the attractive potential energy, the total energy, omitting the translational energy of the centre of mass of the pair, can take only certain discrete values.

The most important correction is the first. It was pointed out by Slater (1931), developed by Uhlenbeck & Gropper (1932), Wigner (1932), and most conveniently by Kirkwood (1933). Uhlenbeck & Beth (1936) have illustrated the subject by application to particular cases.

The second correction, for the effect of statistics, has been shown by Beth & Uhlenbeck (1937) to be negligible except for helium at low temperatures.

The correction for discrete states has been evaluated by Beth & Uhlenbeck (1937) in terms of the phase shifts of the wave functions in the potential field. The correction is appreciable only when the number of discrete levels is small and the temperature is low, both conditions being necessary. Even for helium, which has at most one discrete level,* the correction is appreciable only at low temperatures. For molecules such as argon, with something like a dozen discrete levels, the effect is negligible at all temperatures. To sum up, only the correction (I) need be taken into account except for helium at low temperatures, where it is necessary to include all three corrections.

Thus with the first quantum correction, or rather the first approximation to it,

$$B = 2\pi n\int_0^\infty \{1-e^{-E/kT}\}\, r^2\, dr + \frac{nh^2}{24\pi m(kT)^3}\int_0^\infty \left(\frac{dE}{dr}\right)^2 r^2 e^{-E/kT}\, dr, \tag{4}$$

where m is the mass of the molecule. There are terms involving higher powers of h^2/mkT, but these are appreciable at low temperatures only, and may otherwise be neglected. We reduce (4) to a more convenient form by the following substitutions:

$$\lambda^2 = \frac{h^2}{2\pi mkT}, \quad \sigma = \frac{r}{r_0}, \quad E = -\epsilon f(\sigma), \quad x = \frac{\epsilon}{kT}, \tag{5}$$

where r_0 and ϵ are the position and depth of the minimum of $E(r)$. Hence

$$B = 2\pi n r_0^3\{F_0(x) + (\lambda/r_0)^2\, F_1(x)\}, \tag{6}$$

where

$$F_0(x) = \int_0^\infty \{1-e^{xf}\}\, \sigma^2\, d\sigma \tag{7}$$

and

$$F_1(x) = \frac{x^2}{24\pi}\int_0^\infty \left(\frac{df}{d\sigma}\right)^2 \sigma^2 e^{xf}\, d\sigma. \tag{8}$$

For f, the shape of the potential, we use the following forms:

$$\sigma \geqslant 1, \quad f = f_1\sigma^{-6}(1+\beta\sigma^{-2}) - f_2 e^{-\alpha(\sigma-1)}, \tag{9}$$

$$\sigma < 1, \quad f = f_1\sigma^{-6}(1+\beta\sigma^{-2})\exp\left\{-4\left(\frac{1}{\sigma}-1\right)^3\right\} - f_2 e^{-\alpha(\sigma-1)}, \tag{10}$$

---

* The existence of this level is not yet established with certainty, as it depends critically on the form assumed for the intermolecular potential.

where $$f_1 = \alpha/\{\alpha(1+\beta)-6-8\beta\},$$

$$f_2 = (6+8\beta)/\{\alpha(1+\beta)-6-8\beta\} = -1+(1+\beta)f_1.$$

In (9) there are attractive potentials proportional to r^{-6} and r^{-8}, and an exponential repulsion. The parameter α measures the steepness of the repulsion, and values deduced for particular gases by various authors lie in the neighbourhood of 13·5; our tables cover the range 12·5 to 14·5. The importance of the r^{-8} term, relative to the r^{-6} potential, is measured by β. Margenau (1939) has given a formula for the r^{-8} potential which makes β usually about 0·15; we have used the values $\beta = 0$ and 0·2.

There is a discontinuity in the third derivative of f in passing through $\sigma = 1$. This is due to the multiplication of the attractive potential by a factor which prevents the whole potential going negative and finally infinite as σ tends to zero, and which therefore will make easier a comparison of the potentials with the results of scattering experiments (cf. Amdur & Pearlman 1941).

Whitelaw (1934) has shown that if c_p^0 and μ_0 are respectively the specific heat at constant pressure, per molecule, and the Joule-Thomson coefficient, both extrapolated to zero pressure, then

$$c_p^0\mu_0 = T^2\frac{d}{dT}\left(\frac{B}{nT}\right), \tag{11}$$

and Hirschfelder *et al.* (1938) have used this equation to derive theoretical expressions for μ_0. In the case of the potential (9) and (10) one can write

$$c_p^0\mu_0 = 2\pi r_0^3\{G_0(x)+(\lambda/r_0)^2\,G_1(x)\}, \tag{12}$$

where $$G_0 = \int_0^\infty (e^{xf}+xfe^{xf}-1)\,\sigma^2 d\sigma \tag{13}$$

and $$G_1 = -\frac{x^2}{24\pi}\int_0^\infty \left(\frac{df}{d\sigma}\right)^2 (4+xf)\,e^{xf}\sigma^2 d\sigma. \tag{14}$$

G_0 is the classical term and G_1 the first and most important of the three quantum corrections.

TABLES OF THE FUNCTIONS F_0, F_1, G_0 AND G_1

These tables cover the following ranges: α from 12·5 to 14·5, x from 0·04 to 0·8, and the values of β are 0 and 0·2.

Interpolation with respect to α and x is possible, but with only two values of β it is not possible to interpolate with respect to β, except by a linear interpolation. However, the variation with β is small, and the error in a linear interpolation is probably negligible.

The lower limit of the tabulated values of x, namely, 0·04, was chosen because x is greater than 0·04 for all known virial measurements, except certain results for helium at high temperatures. The tabulated values for $x = 0·04$ and $\beta = 0$ are at present extrapolated. An extension of the tables to this and lower values of x is in progress, however, with a view to their application to the special case of helium.

The upper limit of x, which is 0·8, lies in the neighbourhood of the critical temperature. It is not necessary to carry the tabulation to lower temperatures, because the form of the curves in this region is such that a unique fit to experimental data is impossible. Observations well above the critical temperature are essential to the derivation of an intermolecular potential.

These functions have no special mathematical interest, so the accuracy to be attained was settled solely by considerations of their use in analysis of observed second virial and Joule-Thomson data. These are subject to errors of at least 1 % of the values found at the critical temperature. We have aimed at a maximum error of less than 0·1 % in F_0 and G_0 and 1 % in F_1 and G_1, on this scale, and the intervals used in the computation have been chosen to reduce the errors to these limits.

The tables are all to be interpolated in the same way. The printed differences are of two types: those for interpolation with respect to α are printed between the columns, those for interpolation with respect to x are placed between the rows. The differences are always the sum of adjacent second differences, modified by a 'throw-back' to include the effect of fourth differences. These sums are sufficient for Bessel interpolation:

$$f(a+nh) = nf(a+h)+(1-n)f(a)+B''(n)\,M,$$

where n lies between 0 and 1, $B''(n)$ is the Bessel coefficient of the second difference, and M is the printed difference-sum for the interval $(a, a+h)$. Where M is omitted, linear interpolation introduces an error of less than half a unit of the last place.

COMPUTATION OF THE TABLES

For $\sigma \geqslant 1$, $e^{xf(\sigma)}$ can be expanded as a power series, and if one writes

$$F_{0n} = \int_1^\infty f^n \sigma^2 d\sigma \tag{15}$$

and

$$F_{1n} = \int_1^\infty \left(\frac{df}{d\sigma}\right)^2 f^n \sigma^2 d\sigma, \tag{16}$$

then

$$F_0 = \int_0^1 (1-e^{xf})\,\sigma^2 d\sigma - \sum_{n=1}^\infty \frac{x^n}{n!} F_{0n} \tag{17}$$

and

$$F_1 = \frac{x^2}{24\pi}\int_0^1 \left(\frac{df}{d\sigma}\right)^2 e^{xf}\sigma^2 d\sigma + \frac{x^2}{24\pi}\sum_{n=0}^\infty \frac{x^n}{n!} F_{1n}, \tag{18}$$

with analogous series for G_0 and G_1. The integrals F_{0n} and F_{1n} were calculated for n up to 7. Values of F_0, F_1, G_0 and G_1 were computed for $\alpha = 12{\cdot}5\,(0{\cdot}5)\,14{\cdot}5$ and $x = 0{\cdot}04$, $0{\cdot}05$, $0{\cdot}06$, $0{\cdot}08$, $0{\cdot}1$, $0{\cdot}15$, $0{\cdot}2\,(0{\cdot}1)\,0{\cdot}8$, and the tables were completed by interpolation. Apart from checking against the original computations, the proofs have been differenced, and it is believed that few errors are left in the tables. The formulae and the intervals used in the numerical integrations were of two types, and a number of the integrals were computed by both methods as a check on the accuracy attained.

For permission to publish this work we wish to thank the Chief Scientific Officer, Ministry of Supply.

Table 1. The function F_0, related to the second virial coefficient, for $\beta = 0$

x	$\alpha = 12·5$		13		13·5		14		14·5
0·04	(0·1081)		(0·1122)		(0·1161)		(0·1198)		(0·1234)
0·05	0·1079		0·1121		0·1161		0·1199		0·1236
	−20		−20		−20		−20		−20
0·06	0·1066		0·1109		0·1150		0·1189		0·1226
	−15		−15		−15		−15		−15
0·07	0·1044		0·1088		0·1130		0·1170		0·1207
	−12		−12		−12		−12		−11
0·08	0·1016		0·1061		0·1104		0·1144		0·1183
	−9		−9		−9		−9		−9
0·09	0·0982		0·1029		0·1073		0·1114		0·1153
	−7		−7		−7		−7		−7
0·10	0·0945		0·0993		0·1038		0·1080		0·1120
	−138		−137		−136		−134		−133
0·15	0·0716	−8	0·0771		0·0821		0·0868		0·0913
	−71		−70		−69		−69		−68
0·20	0·0444	−10	0·0506	−9	0·0562	−8	0·0615		0·0664
	−48		−48		−47		−46		−46
0·25	0·0144	−12	0·0213	−10	0·0276	−9	0·0335	−8	0·0389
	−38		−37		−37		−36		−36
0·30	−0·0176	−13	−0·0100	−12	−0·0030	−10	0·0035	−9	0·0094
	−32		−32		−31		−31		−30
0·35	−0·0514	−15	−0·0429	−14	−0·0352	−12	−0·0282	−10	−0·0216
	−29		−29		−28		−28		−27
0·40	−0·0866	−17	−0·0774	−15	−0·0689	−13	−0·0612	−11	−0·0541
	−28		−27		−27		−26		−26
0·45	−0·1233	−19	−0·1132	−17	−0·1040	−14	−0·0956	−12	−0·0879
	−27		−26		−26		−25		−25
0·50	−0·1614	−22	−0·1504	−19	−0·1404	−16	−0·1313	−13	−0·1229
	−27		−26		−26		−25		−25
0·55	−0·2008	−24	−0·1889	−21	−0·1781	−17	−0·1682	−13	−0·1592
	−27		−26		−25		−25		−24
0·60	−0·2413	−27	−0·2287	−23	−0·2170	−19	−0·2064	−16	−0·1967
	−27		−26		−26		−25		−25
0·65	−0·2836	−28	−0·2697	−24	−0·2572	−20	−0·2458	−17	−0·2354
	−27		−27		−26		−26		−25
0·70	−0·3270	−30	−0·3121	−26	−0·2987	−22	−0·2865	−19	−0·2753
	−28		−27		−27		−26		−26
0·75	−0·3718	−32	−0·3559	−28	−0·3416	−24	−0·3285	−21	−0·3165
	−28		−28		−27		−27		−26
0·80	−0·4180	−34	−0·4010	−30	−0·3857	−26	−0·3718	−23	−0·3590

TABLE 2. THE FUNCTION F_0, RELATED TO THE SECOND VIRIAL COEFFICIENT, FOR $\beta = 0{\cdot}2$

x	$\alpha = 12{\cdot}5$		13		13·5		14		14·5
0·04	0·1105		0·1147		0·1186		0·1224		0·1261
	(−31)		(−31)		(−31)		(−31)		(−31)
0·05	0·1106		0·1149		0·1189		0·1228		0·1265
	−20		−21		−20		−20		−21
0·06	0·1094		0·1138		0·1180		0·1220		0·1257
	−15		−15		−15		−15		−15
0·07	0·1074		0·1120		0·1163		0·1203		0·1242
	−12		−12		−12		−12		−12
0·08	0·1047		0·1094		0·1138		0·1180		0·1219
	−9		−9		−9		−9		−9
0·09	0·1016		0·1064		0·1109		0·1152		0·1192
	−8		−8		−8		−7		−7
0·10	0·0981		0·1030		0·1076		0·1120		0·1160
	−136		−136		−135		−134		−132
0·15	0·0762	−10	0·0819	−9	0·0871		0·0919		0·0964
	−70		−70		−69		−68		−67
0·20	0·0501	−12	0·0565	−10	0·0623	−8	0·0677		0·0727
	−48		−47		−46		−46		−47
0·25	0·0212	−14	0·0284	−12	0·0349	−10	0·0409	−8	0·0464
	−38		−38		−37		−36		−36
0·30	−0·0097	−16	−0·0017	−14	0·0055	−11	0·0121	−9	0·0182
	−31		−31		−30		−30		−29
0·35	−0·0424	−17	−0·0336	−15	−0·0256	−13	−0·0183	−11	−0·0116
	−29		−28		−28		−27		−27
0·40	−0·0764	−19	−0·0668	−17	−0·0580	−14	−0·0501	−12	−0·0427
	−28		−27		−27		−26		−26
0·45	−0·1119	−20	−0·1014	−18	−0·0919	−16	−0·0832	−14	−0·0753
	−26		−26		−25		−25		−24
0·50	−0·1488	−22	−0·1373	−20	−0·1269	−18	−0·1175	−16	−0·1089
	−26		−26		−25		−25		−24
0·55	−0·1869	−24	−0·1745	−22	−0·1633	−20	−0·1531	−17	−0·1439
	−26		−25		−25		−24		−24
0·60	−0·2264	−27	−0·2130	−24	−0·2009	−21	−0·1899	−18	−0·1799
	−26		−26		−25		−25		−24
0·65	−0·2672	−30	−0·2527	−27	−0·2397	−23	−0·2280	−19	−0·2172
	−26		−26		−26		−25		−25
0·70	−0·3093	−33	−0·2938	−29	−0·2798	−25	−0·2672	−20	−0·2557
	−27		−27		−26		−26		−25
0·75	−0·3527	−37	−0·3361	−32	−0·3212	−26	−0·3078	−21	−0·2955
	−27		−27		−27		−27		−26
0·80	−0·3975	−40	−0·3798	−34	−0·3639	−28	−0·3496	−22	−0·3365

TABLE 3. THE FUNCTION F_1, RELATED TO THE QUANTUM CORRECTION TO THE SECOND VIRIAL COEFFICIENT, FOR $\beta = 0$

x	$\alpha = 12·5$	13	13·5	14	14·5
0·04	(0·100)	(0·104)	(0·109)	(0·114)	(0·119)
0·05	0·107	0·111	0·116	0·121	0·127
0·06	0·113	0·118	0·123	0·128	0·134
0·07	0·119	0·124	0·129	0·134	0·140
0·08	0·124	0·129	0·135	0·140	0·146
0·09	0·129	0·135	0·140	0·146	0·152
0·10	0·134	0·140	0·145	0·151	0·157
0·15	0·157	0·164	0·170	0·176	0·183
0·20	0·179	0·186	0·193	0·200	0·208
0·25	0·201	0·209	0·217	0·225	0·233
0·30	0·224	0·232	0·241	0·250	0·258
0·35	0·247	0·257	0·266	0·276	0·285
0·40	0·272	0·282	0·293	0·303	0·314
0·45	0·299	0·310	0·321	0·332	0·344
0·50	0·326	0·339	0·351	0·363	0·376
0·55	0·356	0·370	0·383	0·396	0·410
0·60	0·388	0·402	0·417	0·432	0·447
0·65	0·422	0·438	0·453	0·470	0·486
0·70	0·458	0·475	0·493	0·510	0·528
0·75	0·497	0·515	0·534	0·554	0·573
0·80	0·539	0·559	0·579	0·600	0·621

TABLE 4. THE FUNCTION F_1, RELATED TO THE QUANTUM CORRECTION TO THE SECOND VIRIAL COEFFICIENT, FOR $\beta = 0·2$

x	$\alpha = 12·5$	13	13·5	14	14·5
0·04	0·103	0·107	0·112	0·117	0·122
0·05	0·110	0·114	0·119	0·124	0·129
0·06	0·116	0·121	0·126	0·131	0·136
0·07	0·122	0·127	0·132	0·137	0·143
0·08	0·127	0·132	0·138	0·143	0·149
0·09	0·132	0·137	0·143	0·149	0·155
0·10	0·137	0·142	0·148	0·154	0·160
0·15	0·160	0·166	0·172	0·179	0·186
0·20	0·182	0·189	0·196	0·203	0·210
0·25	0·203	0·211	0·219	0·227	0·234
0·30	0·225	0·234	0·242	0·251	0·259
0·35	0·248	0·257	0·266	0·275	0·285
0·40	0·272	0·282	0·292	0·302	0·312
0·45	0·296	0·307	0·318	0·329	0·340
0·50	0·323	0·335	0·346	0·358	0·370
0·55	0·351	0·363	0·376	0·388	0·401
0·60	0·380	0·394	0·407	0·421	0·434
0·65	0·411	0·426	0·441	0·455	0·469
0·70	0·445	0·460	0·476	0·491	0·506
0·75	0·481	0·496	0·513	0·529	0·545
0·80	0·518	0·536	0·553	0·571	0·588

Table 5. The function G_0, related to the Joule-Thomson coefficient, for $\beta = 0$

x	$\alpha = 12·5$		13		13·5		14		14·5
0·04	(−0·1108)		(−0·1153)		(−0·1196)		(−0·1236)		(−0·1273)
0·05	−0·1040		−0·1087		−0·1132		−0·1173		−0·1213
	23		23		23		23		23
0·06	−0·0958	8	−0·1008		−0·1054		−0·1098		−0·1139
	17		17		17		17		16
0·07	−0·0866	8	−0·0919	8	−0·0967		−0·1013		−0·1055
	13		13		13		13		13
0·08	−0·0767	9	−0·0822	8	−0·0873		−0·0920		−0·0965
	10		10		10		10		10
0·09	−0·0662	10	−0·0720	9	−0·0773		−0·0822		−0·0869
	8		8		8		8		8
0·10	−0·0552	11	−0·0613	9	−0·0669	8	−0·0720		−0·0768
	172		169		167		165		163
0·15	+0·0047	14	−0·0028	12	−0·0096	10	−0·0159	8	−0·0217
	104		102		101		99		97
0·20	0·0706	18	0·0616	15	0·0534	13	0·0459	11	0·0390
	84		82		80		79		77
0·25	0·1409	21	0·1303	18	0·1207	16	0·1119	13	0·1038
	76		74		72		71		70
0·30	0·2152	25	0·2029	21	0·1917	18	0·1816	15	0·1723
	72		71		69		68		67
0·35	0·2931	29	0·2790	25	0·2663	21	0·2547	18	0·2441
	72		71		69		68		66
0·40	0·3747	32	0·3587	29	0·3442	25	0·3311	21	0·3192
	74		72		71		69		67
0·45	0·4598	38	0·4419	33	0·4257	28	0·4110	23	0·3976
	76		74		72		71		69
0·50	0·5487	42	0·5287	36	0·5107	31	0·4944	25	0·4794
	78		76		75		73		72
0·55	0·6415	47	0·6193	40	0·5994	34	0·5813	29	0·5648
	81		79		77		76		75
0·60	0·7382	52	0·7138	44	0·6919	38	0·6720	32	0·6538
	85		83		81		79		78
0·65	0·8391	57	0·8124	49	0·7883	41	0·7665	34	0·7466
	89		86		84		83		81
0·70	0·9443	63	0·9152	54	0·8889	45	0·8651	37	0·8434
	93		90		88		86		85
0·75	1·0541	67	1·0224	59	0·9938	50	0·9679	40	0·9443
	97		95		92		90		89
0·80	1·1686	72	1·1342	64	1·1031	55	1·0752	44	1·0496

TABLE 6. THE FUNCTION G_0, RELATED TO THE JOULE-THOMSON COEFFICIENT, FOR $\beta = 0·2$

x	$\alpha = 12·5$		13		13·5		14		14·5
0·04	−0·1139		−0·1186		−0·1230		−0·1271		−0·1309
	33		33		32		31		31
0·05	−0·1075		−0·1124		−0·1170		−0·1213		−0·1253
	23		23		23		22		22
0·06	−0·0997	8	−0·1049		−0·1097		−0·1142		−0·1183
	17		17		17		17		17
0·07	−0·0909	9	−0·0964	8	−0·1014		−0·1061		−0·1105
	13		13		13		13		13
0·08	−0·0814	10	−0·0872	9	−0·0924		−0·0973		−0·1018
	10		10		10		10		10
0·09	−0·0714	10	−0·0774	9	−0·0829	8	−0·0880		−0·0927
	9		9		8		8		8
0·10	−0·0608	11	−0·0671	10	−0·0729	9	−0·0782	8	−0·0831
	170		168		166		163		161
0·15	−0·0031	17	−0·0109	14	−0·0180	11	−0·0244	9	−0·0304
	104		102		100		98		96
0·20	+0·0606	20	+0·0512	17	+0·0427	14	+0·0350	11	+0·0279
	82		80		78		77		75
0·25	0·1287	24	0·1176	20	0·1076	17	0·0985	14	0·0903
	74		72		70		69		68
0·30	0·2005	28	0·1876	24	0·1761	20	0·1656	17	0·1560
	72		70		68		67		66
0·35	0·2760	32	0·2613	28	0·2480	24	0·2361	20	0·2252
	71		69		68		66		65
0·40	0·3550	37	0·3383	32	0·3233	27	0·3098	22	0·2975
	72		71		69		67		66
0·45	0·4376	42	0·4188	36	0·4020	30	0·3869	24	0·3731
	74		72		70		69		68
0·50	0·5238	47	0·5029	40	0·4842	33	0·4674	27	0·4520
	77		75		73		71		70
0·55	0·6138	51	0·5907	44	0·5700	37	0·5513	30	0·5344
	80		78		76		74		73
0·60	0·7077	56	0·6823	48	0·6595	41	0·6389	33	0·6203
	83		81		79		77		76
0·65	0·8057	62	0·7778	53	0·7529	44	0·7304	35	0·7099
	86		84		83		81		79
0·70	0·9080	68	0·8775	58	0·8503	48	0·8257	38	0·8034
	91		89		86		84		83
0·75	1·0146	74	0·9815	63	0·9520	52	0·9252	41	0·9009
	95		92		89		87		85
0·80	1·1260	80	1·0901	69	1·0580	58	1·0290	46	1·0027

TABLE 7. THE FUNCTION $-G_1$, RELATED TO THE QUANTUM CORRECTION TO THE JOULE-THOMSON COEFFICIENT, FOR $\beta = 0$

x	$\alpha = 12{\cdot}5$	13	13·5	14	14·5
0·04	(0·229)	(0·239)	(0·249)	(0·259)	(0·270)
0·05	0·247	0·257	0·267	0·278	0·290
0·06	0·262	0·273	0·284	0·295	0·307
0·07	0·277	0·288	0·299	0·311	0·323
0·08	0·291	0·302	0·314	0·326	0·339
0·09	0·304	0·316	0·328	0·341	0·354
0·10	0·317	0·330	0·342	0·355	0·369
0·15	0·381	0·396	0·411	0·426	0·441
0·20	0·446	0·463	0·480	0·497	0·514
	8	8	9	9	10
0·25	0·514	0·533	0·553	0·572	0·592
	11	11	11	11	12
0·30	0·586	0·608	0·630	0·652	0·674
	12	13	13	14	15
0·35	0·664	0·689	0·714	0·739	0·764
	14	14	15	16	16
0·40	0·749	0·776	0·805	0·833	0·861
	15	16	17	18	19
0·45	0·841	0·872	0·903	0·935	0·967
	17	18	19	20	21
0·50	0·941	0·976	1·011	1·047	1·083
	19	20	21	22	23
0·55	1·050	1·089	1·129	1·169	1·210
	21	22	23	24	25
0·60	1·169	1·213	1·257	1·303	1·349
	23	24	25	26	27
0·65	1·298	1·348	1·398	1·449	1·500
	24	25	26	27	28
0·70	1·440	1·495	1·551	1·609	1·665
	26	27	28	29	29
0·75	1·595	1·656	1·718	1·782	1·845
0·80	1·762	1·830	1·900	1·971	2·040

TABLE 8. THE FUNCTION $-G_1$, RELATED TO THE QUANTUM CORRECTION TO THE JOULE-THOMSON COEFFICIENT, FOR $\beta = 0{\cdot}2$

x	$\alpha = 12{\cdot}5$	13	13·5	14	14·5
0·04	0·235	0·245	0·255	0·265	0·276
0·05	0·252	0·262	0·273	0·284	0·295
0·06	0·268	0·278	0·289	0·301	0·312
0·07	0·282	0·293	0·305	0·316	0·329
0·08	0·296	0·307	0·319	0·332	0·344
0·09	0·309	0·321	0·334	0·346	0·359
0·10	0·322	0·335	0·348	0·361	0·374
0·15	0·385	0·400	0·415	0·430	0·445
0·20	0·449	0·466	0·483	0·500	0·517
0·25	0·515	0·534	0·553	0·572	0·591
	8	8	8	9	10
0·30	0·584	0·606	0·627	0·648	0·669
	10	10	11	12	12
0·35	0·659	0·682	0·706	0·729	0·752
	11	12	13	13	14
0·40	0·738	0·764	0·791	0·817	0·842
	13	14	14	14	15
0·45	0·824	0·853	0·882	0·911	0·940
	14	15	15	15	15
0·50	0·916	0·949	0·981	1·013	1·045
	16	16	17	16	16
0·55	1·016	1·052	1·088	1·123	1·157
	18	17	18	18	18
0·60	1·125	1·164	1·203	1·241	1·277
	19	19	19	19	20
0·65	1·242	1·285	1·327	1·368	1·408
	20	20	21	21	22
0·70	1·370	1·416	1·461	1·506	1·549
	20	21	22	23	24
0·75	1·507	1·557	1·606	1·654	1·702
	21	22	22	23	24
0·80	1·654	1·709	1·762	1·815	1·867

REFERENCES

Amdur, I. & Pearlman, H. 1941 *J. Chem. Phys.* **9**, 503.
Beth, E. & Uhlenbeck, G. E. 1937 *Physica, 's Grav.* **4**, 915.
Bleick, W. E. & Mayer, J. E. 1934 *J. Chem. Phys.* **2**, 252.
Born, M. & Mayer, J. E. 1932 *Z. Phys.* **75**, 1.
Buckingham, R. A. 1938 *Proc. Roy. Soc.* A, **168**, 264.
Fowler, R. H. 1936 *Statistical mechanics*, chapter x. Camb. Univ. Press.
Herzfeld, K. F. 1937 *Phys. Rev.* **52**, 374.
Hirschfelder, J. O., Ewell, R. B. & Roebuck, J. R. 1938 *J. Chem. Phys.* **6**, 205.
Huggins, M. L. 1937 *J. Chem. Phys.* **5**, 143.
Kirkwood, J. G. 1933 *Phys. Rev.* **44**, 31.
Lennard-Jones, J. E. 1924 *Proc. Roy. Soc.* A, **106**, 463.
Margenau, H. 1939 *Rev. Mod. Phys.* **11**, 1.
Orr, W. J. C. 1942 *Proc. Camb. Phil. Soc.* 38, 224.
Slater, J. C. 1931 *Phys. Rev.* **38**, 237.
Uhlenbeck, G. E. & Beth, E. 1936 *Physica, 's Grav.* **3**, 729.
Uhlenbeck, G. E. & Gropper, L. 1932 *Phys. Rev.* **41**, 79.
Whitelaw, N. G. 1934 *Physica, 's Grav.* **1**, 749.
Wigner, E. 1932 *Phys. Rev.* **40**, 749.

The oblique reflexion of very long wireless waves from the ionosphere

By M. V. Wilkes, *Mathematical Laboratory, Cambridge*

(*Communicated by W. V. D. Hodge, F.R.S.—Received* 23 *July* 1946)

An attempt is made in the paper to provide a satisfactory theoretical basis for a future discussion of the experimental data on the propagation of very long radio waves (18,800 m.) given by Best, Ratcliffe & Wilkes, and Budden, Ratcliffe & Wilkes. The reflexion of very long plane waves incident obliquely on a horizontally stratified ionized medium with a vertical magnetic field is first considered in general terms, and it is shown that the medium can be divided into a transition region and a reflecting region.

If the ionization in the reflecting region increases linearly with height it is shown that propagation is governed by the following equations:

$$\frac{\partial^2 L}{\partial \zeta^2}+(\alpha+\zeta)\,L+\beta M=0,\quad \frac{\partial^2 M}{\partial \zeta^2}+(\alpha-\zeta)\,M+\beta L=0,$$

where α and β are constants depending on the angle of incidence. Under the conditions of the experiments β is small, and a solution, in terms of contour integrals, valid in this case is obtained.

1. Introduction

During the period 1934–9 a series of observations were made at Cambridge on the nature of the downcoming wave received from the Post Office transmitter at Rugby on a wave-length of 18,800 m. (Best, Ratcliffe & Wilkes 1936; Budden, Ratcliffe & Wilkes 1939). These observations were supplemented by measurements made with a mobile field strength measuring set at a large number of points at distances between 70 and 380 km. from the transmitter. The experiments at Cambridge corresponded to an angle of incidence of the reflected wave on the ionosphere of about 30°, and those with the mobile apparatus to angles between about 25° and 70°.

The theoretical interpretation of these experimental results presents some difficulty, arising from the fact that with very long waves it is not possible to make the assumption which is admissible in the case of short waves, that the parameters of the ionosphere change by only a small fraction of themselves in a distance equal to the wave-length (Booker 1936, 1938; Försterling & Lassen 1933). In other words, the problem may not be treated by the methods of geometrical optics, and any theoretical discussion must proceed from a consideration of the differential equations of the magneto-ionic theory.

Only in the case of vertical incidence (with the further assumption that the earth's magnetic field is vertical) is it found that these equations can be handled at all easily, and this case has been discussed in detail by the present author (Wilkes 1940). There are then two components which are propagated independently in the ionosphere. The experiments indicated, as far as they went, that the character of the reflected wave did not depend markedly on the angle of incidence, and it was therefore thought that some understanding of the mechanism of reflexion might be obtained by applying vertical incidence theory to the case of waves received at

Cambridge, after reflexion at an angle of incidence of about 30°. While this method undoubtedly leads to a general explanation of the observed phenomena, it cannot be regarded as logically satisfactory until it is shown that there is no large change in the nature of the reflected wave as the angle of incidence is reduced from a few degrees to zero; there is the possibility that the radical simplification of the equations which takes place when incidence becomes normal may be accompanied by a radical change in the mechanism of reflexion. Further, the vertical incidence theory can of necessity throw no light on the variation of the phase of the reflected wave with angle of incidence. This information is needed in order to enable the true height of reflexion to be deduced from the data given in figure 1 of Budden *et al.* (1939); at present only an effective height can be obtained.

An attempt will be made in this paper to give a more satisfactory basis for the theoretical interpretation of measurements made at short and medium distances from a very long wave transmitter, by discussing the problem of reflexion at oblique incidence. The method of approach will be that of direct solution of the field equations subject to the condition that the wave-length is very long. The assumption that the earth's magnetic field is vertical will be retained.

It will be found that the solution obtained consists of the solution for the case of vertical incidence, together with certain correcting terms, which tend to zero as the angle of incidence is reduced. The use of the vertical incidence results as an approximation to the solution in the case of oblique incidence is in this way justified.

In § 2 the equations of the magneto-ionic theory will be written down, and in § 3 it will be shown that the ionosphere may be divided into a 'transition' region and a 'reflecting' region, of which the latter is the more important from the present point of view. In § 4 (*a*) the solution of the equations of propagation in the reflecting region, assuming a linear variation of ionization with height, will be shown to depend on the solution of a certain second-order ordinary differential equation, and in § 4 (*b*) a solution of this equation in terms of a series of functions will be obtained. Asymptotic forms will be given in § 4 (*c*).

Where it is necessary to consider numerical values, those appropriate to the experiments described by Best *et al.* (1936) and Budden *et al.* (1939) will be taken. Detailed discussion of the application of the analysis to the interpretation of these experiments will, however, be reserved for another paper. A brief indication of the lines this discussion will follow will be given in § 5.

2. THE FIELD EQUATIONS

The following notation will be used:

c = velocity of electromagnetic waves *in vacuo*.

m = mass of an electron.

e = charge of an electron.

N = electron density.

$\mathbf{H}_e$ = the earth's magnetic field assumed vertical.

$p = 2\pi \times$ frequency of waves.

ν = a quantity which determines the effective damping, and is proportional to the frequency of electronic collisions.

$X = \dfrac{Ne^2}{mp^2\epsilon_0}$.*

$\mathbf{Y} = \dfrac{e\mathbf{H}_e\mu_0}{mp}$.

$Y = |\mathbf{Y}|$.*

$Z = \nu/p$.*

$\mathbf{E}$ = electric field of the wave.

$\mathbf{H}$ = magnetic field of the wave.

$\mathbf{P}$ = polarization of the medium.

ϵ_0 = dielectric constant of free space ($8{\cdot}854\,pF/m$ in practical units).

μ_0 = permeability of free space ($4\pi \times 10^{-7}$ henries/m. in practical units).

$c = (\mu_0\epsilon_0)^{-\frac{1}{2}}$ = velocity of electromagnetic waves *in vacuo*.

$k = p/c$.

H = scale height of the atmosphere.

The field equations are

$$\operatorname{curl}\mathbf{E} = -\mu_0\frac{\partial \mathbf{H}}{\partial t}, \quad \operatorname{curl}\mathbf{H} = \epsilon_0\frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t}.$$

These lead to

$$\begin{aligned}\operatorname{curl}\operatorname{curl}\mathbf{E} &= \mu_0\epsilon_0 p^2\mathbf{E} + p^2\mu_0\mathbf{P}\\ &= k^2\mathbf{E} + p^2\mu_0\mathbf{P},\end{aligned} \tag{2·1}$$

where a factor e^{ipt} has been taken.

The polarization $\mathbf{P}$ is thought of as arising from the displacement of the electrons in the medium under the action of the applied electromagnetic field of the wave from the orbits they would otherwise describe. As is usual in developing the magneto-ionic theory, it is assumed that this treatment is applicable to a medium containing free electrons. The extra polarization term of the Lorentz theory will not be included. If the magnetic field of the wave is neglected it follows that

$$\frac{\partial^2\mathbf{P}}{\partial t^2} + \nu\frac{\partial \mathbf{P}}{\partial t} = \frac{Ne^2}{m}\mathbf{E} + \frac{\mu_0 e}{m}\mathbf{H}_e \wedge \frac{\partial \mathbf{P}}{\partial t}$$

or

$$\mathbf{P}(1 - iZ) = -X\epsilon_0\mathbf{E} - i\mathbf{Y}\wedge\mathbf{P}. \tag{2·2}$$

---

* These three quantities are often denoted by x, y, z. Large letters have been used in order to avoid confusion with the spacial co-ordinates.

Since this analysis will ultimately be applied to the case of a plane wave incident at an angle θ on a stratified medium, it is convenient to assume that all the dependent variables are proportional to $e^{ik\sin\theta x}$ and are independent of y. The variation in the x direction will then be the same as that in the plane wave in free space, and Snell's law will be automatically satisfied.† If the components of the vectors **E** and **P** divided by this factor are denoted by E_x, E_y, E_z, and P_x, P_y, P_z, it follows, on expanding (2·1), that

$$\left.\begin{aligned} \frac{d^2E_x}{dz^2}-ik\sin\theta\frac{dE_z}{dz} &= -k^2E_x-p^2\mu_0P_x,\\ \frac{d^2E_y}{dz^2}-k^2\sin^2\theta\,E_y &= -k^2E_y-p^2\mu_0P_y,\\ -ik\sin\theta\frac{dE_x}{dz}-k^2\sin^2\theta\,E_z &= -k^2E_z-p^2\mu_0P_z. \end{aligned}\right\} \tag{2·3}$$

Expanding (2·2), and solving for P_x, P_y, P_z, taking the magnetic field vertical, we have

$$P_x = -\frac{X\epsilon_0}{(1-iZ)^2-Y^2}[(1-iZ)\,E_x+iY\,E_y],$$

$$P_y = \frac{X\epsilon_0}{(1-iZ)^2-Y^2}[iY\,E_x-(1-iZ)\,E_y],$$

$$P_z = -\frac{X\epsilon_0}{1-iZ}E_z.$$

Finally, substituting these values in (2·3) and rearranging, we have

$$\frac{d^2}{dz^2}(E_x\pm iE_y)+\frac{d}{dz}\left(\frac{\sin^2\theta}{\cos^2\theta-X/(1-iZ)}\frac{dE_x}{dz}\right)$$
$$= -k^2(E_x\pm iE_y)\left[1-\frac{X}{1\mp Y-iZ}\right]\pm ik^2\sin^2\theta\,E_y, \tag{2·4}$$

$$E_z = \frac{i}{k}\frac{\sin\theta}{\cos^2\theta-X/(1-iZ)}\frac{dE_x}{dz}. \tag{2·5}$$

3. THE TRANSITION AND REFLECTING REGIONS

It is important to have an appreciation of the numerical values assumed by the quantities X and Y in the practical cases presented here. Y may be expressed as the ratio of the gyromagnetic frequency of electrons in the field of the earth to the frequency of the waves, and consequently becomes large for low frequencies. For waves of frequency 16 kyc./sec. and latitude 52° the numerical value of Y is about 80, if the magnetic field concerned is taken to be the vertical component of the earth's magnetic field. The expressions $1 \pm Y$ in (2·4) may therefore, without sensible error, be replaced by $\pm Y$.

---

† An ordinary transmitting aerial may be supposed to radiate plane waves in all directions, and in order to discuss the characteristics of the wave received at a fixed point after reflexion by the stratified medium, the value of θ corresponding to the predominant group must be determined by the principle of stationary phase (Hartree 1931; Havelock 1914).

X is proportional to the ionization density, N, and if there is no imposed magnetic field, and no collisional absorption, it is well known that the refractive index q is given by

$$q^2 = 1 - X.$$

The reflecting level for short waves is therefore—in the case of vertical incidence—given by $X = 1$. If a magnetic field is introduced, it is found that the level $X = 1$ is still of importance, but that there is now another level at which reflexion may take place, namely, where $X = 1 + Y$ (Booker 1934, 1935; Taylor 1934). (It is assumed that the frequency is less than the gyromagnetic frequency.) It might be expected that these two levels would retain their importance in the case of very long waves, even though the application of wave theory, rather than geometrical optics, is required for the solution of the problem. If, however, one puts numerical values corresponding to a frequency of 16 kyc./sec. in the equation defining X, one finds, approximately, $X = \frac{1}{3}N$. The level $X = 1$ thus corresponds to an ionization density of only 3 electrons per c.c., and there is no doubt that such a density will be reached at very low levels in the atmosphere where the electrons are so heavily damped by collisions that they are without effect on the passage of the waves. The level $X = 1$ (or $X = \cos^2\theta$, in the case of oblique incidence at an angle θ) thus loses its significance in the case of very long waves, and attention is to be concentrated on the higher level where $X = 1 + Y$ ($= 80$ approximately).

The maximum daytime ionization in the E region is of the order of several times 10^5, so that the maximum value of X is of the order of 10^5. Viewed from the point of view of reflecting very long waves, the ionosphere is thus highly over-dense.

The changes which are observed to occur in the constitution of the ionosphere as the height of the level under consideration is increased will now be examined with special reference to the magnitude of the second term in equation (2·4), which may be written

$$\frac{d}{dz}\left(l\frac{dE_x}{dz}\right), \quad \text{where} \quad l = \frac{\sin^2\theta}{\cos^2\theta - X/(1-iZ)}. \tag{3·1}$$

The frequency of electronic collisions decreases with increasing height in proportion to the decrease of air pressure. In a region where the scale height of the atmosphere may be regarded as constant, Z will therefore decrease exponentially with height according to the equation

$$Z = Z_0 e^{-z/H},$$

where Z_0 is a constant.

The electron density, on the other hand, is an increasing function of height in the region where very long waves are reflected. If it is assumed that reflexion takes place at the bottom of a region of the type described by Chapman (1931*a*, *b*, 1939), the variation of X with height may be expressed by the equation

$$X = X_0 \exp\left[\tfrac{1}{2}(1 - z/H - e^{-z/H} f(\chi))\right],$$

where the height z is measured from a certain reference level, and where $f(\chi)$ is a function of the zenith angle χ of the sun which approximates to sec χ for small values of χ.

Figure 1 has been drawn to illustrate the variation of Z and X with height. The value of X_0 taken is 10^5, and the scale height 6 km., the value deduced from the experiments described in Budden *et al.* (1939). As the zenith angle of the sun changes, the curve for X moves up and down the height axis by an amount approximately equal to the change in $\log f(\chi)$ multiplied by the scale height. The curve for Z has been drawn with an arbitrary value of Z_0; any alteration in the value of this constant will have the effect of moving the curve bodily up or down the height axis.

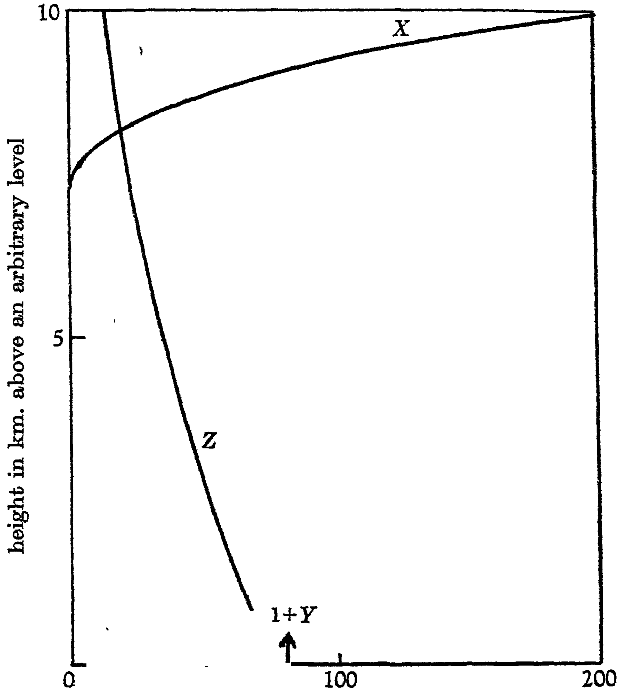


FIGURE 1

At very low levels, l is real and equal to $\tan^2\theta$, and dl/dz is zero. As the height increases l becomes complex, and its modulus may increase in value, depending on the value of Z_0. At high levels, Z decreases to a small value, and X becomes large; both l and dl/dz become small compared with unity and k respectively, and the second term on the left-hand side of equation (2·4) may therefore be neglected. The region in which this approximation may be made, and in which at the same time Z is negligibly small compared with Y, will be referred to as the *reflecting region*, and the region immediately below it as the *transition region*. The latter extends down to levels where propagation is no longer influenced by the presence of ionization. The use of these terms is justified below, where it is shown that on the wave-lengths concerned no reflexion of energy takes place below the top of the transition region.

The height at which the second term in equation (2·4) first becomes negligible may easily be found for the case of a Chapman region. From equation (3·1), if Z is put equal to zero, and X assumed to be large, it follows that

$$\begin{aligned}\frac{dl}{dz} &= \frac{\sin^2\theta}{X^2}\frac{dX}{dz}\ \text{(approx.)}\\ &= \frac{\sin^2\theta}{HX}[\log(X_0/X)-\tfrac{1}{2}z/H].\end{aligned}$$

Taking for purposes of illustration the case referred to above in which the wavelength is 18·8 km., we have the following numerical values; X_0 is taken equal to 10^5, H equal to 6 km. and $\chi = 0$:

X	z km.	l	$\frac{1}{k}\frac{dl}{dz}$
5	$-19{\cdot}1$	$-0{\cdot}20\sin^2\theta$	$1{\cdot}15\sin^2\theta$
10	$-18{\cdot}7$	$-0{\cdot}15\sin^2\theta$	$0{\cdot}54\sin^2\theta$
20	$-18{\cdot}3$	$-0{\cdot}05\sin^2\theta$	$0{\cdot}25\sin^2\theta$
80	$-17{\cdot}4$	$-0{\cdot}012\sin^2\theta$	$0{\cdot}054\sin^2\theta$

It will be seen that in this case l and $\frac{1}{k}\frac{dl}{dz}$ are sufficiently small compared with unity to be neglected at all levels above that for which X is greater than about 10 or 20 depending on the value of θ.

Certain conclusions may be drawn about conditions in the transition region. If Z were sufficiently small, l would become very large at the level where $X = \cos^2\theta$, and reflexion of a large part of the incident energy might be expected to occur. As pointed out above, however, consideration of the numerical values involved allows this possibility to be ruled out absolutely, for it is found that the ionization density required would be only one or two electrons per c.c., and there can be no doubt that such a degree of ionization is reached at levels far below that at which Z can be neglected.

A sharp decrease of Z occurs in the transition region, and if X were at all large this would lead to a sudden change in propagation conditions, and consequent reflexion of a considerable part of the energy. It is, however, easy to see that this is not the state of affairs in the ionosphere at any rate as far as waves of length 18·8 km. are concerned, since if the values of Z and X were suitably adjusted to give the observed reflected wave at midday, say, this condition would no longer hold when the curve of X moved upwards with increase in the zenith angle of the sun; in other words, the changes in reflexion height would be accompanied by large changes in polarization and amplitude of the reflected wave, which is the contrary of what is observed (Best *et al.* 1936; Budden *et al.* 1939). One must therefore conclude that Z has fallen away sensibly to zero while X is still quite small. This point is dealt with more fully by Wilkes (1940).

It is not necessarily true that the curve for X follows the shape of a Chapman region down to the lowest levels of the transition region; there may well be departures from it, due, perhaps, to the overlapping of other regions formed by the ionization of different constituents of the atmosphere under the influence of different parts of the sun's spectrum. The fact that in summer some absorption is observed which remains constant while large changes are taking place in the reflexion height, would point to the existence, at some level where Z is not negligible, of a stratum of ionization not depending much on the zenith angle of the sun until near sunset.

With this exception (which only affects the amplitude of the wave, and not its polarization) it will be seen that the evidence is that the lower part of the transition region, i.e. the part in which Z is not negligible, has no appreciable effect on the waves passing through it. The possibility remains, however, that the upper part of the region, where Z is negligible, may produce some change in the polarization, the amount of which will depend on the thickness to be traversed, and on the rate of change of electron density.

The reflecting region is simpler to study than the transition region not only because the second term in equation (2·4) is negligible, but also because only the variation of X with height has to be taken into account, the value of Z being effectively zero. A study of the transition region, on the other hand, involves detailed assumptions about the relative magnitudes of Z and X. Since it is possible, although by no means certain, that the transition region has an inappreciable effect on the waves passing through it, the most profitable way of proceeding appears to be to defer its study until the reflecting region has been studied in detail. It may then be possible, by making use of the experimental data referred to earlier, either to confirm the assumption that the action of the transition region is negligible, or alternatively to estimate its effect and thereby obtain information about the degree of ionization and the collisional frequency in it.

The remainder of this paper will be devoted to the study of the reflecting region with the assumption that ionization varies linearly with height. The analysis is somewhat involved, and the present paper goes no further than to obtain the formal solution of equation (2·4). This solution will involve four arbitrary constants. In order to obtain the phase and amplitude of the reflected wave it is necessary to choose these constants so that the solution is finite at high level and satisfies certain conditions at the lower boundary of the ionosphere, where the solution must join continuously on to one representing an upgoing incident wave and a downgoing reflected wave in free space. It is hoped to deal with these aspects of the problem in a later paper.

4. THE REFLECTING REGION

(a) *Transformation of the equations*

In this region where l is, by definition, negligible, it is convenient to put

$$L = E_x + iE_y, \qquad (4\cdot1)$$

$$M = E_x - iE_y. \qquad (4\cdot2)$$

When l is neglected, and Z put equal to zero, equations (2·4) become

$$\frac{d^2L}{dz^2}+k^2(1-\tfrac{1}{2}\sin^2\theta+X/Y)\,L+\tfrac{1}{2}k^2\sin^2\theta M = 0, \tag{4·3}$$

$$\frac{d^2M}{dz^2}+k^2(1-\tfrac{1}{2}\sin^2\theta-X/Y)\,M+\tfrac{1}{2}k^2\sin^2\theta L = 0, \tag{4·4}$$

where the further approximation has been made of replacing $1\pm Y$ by $\pm Y$.

Before the equations can be solved, it is necessary to assume some law of variation of X with height. In this paper the simplest possible law will be taken, namely,

$$X = KYz, \tag{4·5}$$

where K is a constant. For many purposes this may be taken as a sufficiently close representation of the ionosphere, but if necessary an approximation to the solution in the general case may be obtained by dividing the region into a series of subregions, in each of which the variation can be represented with sufficient precision by a straight line.

Substituting the value of X given by equation (4·5) into equations (4·3) and (4·4), and writing

$$\zeta = (k^2K)^{\frac{1}{3}}z, \tag{4·6}$$

$$\alpha = (1-\tfrac{1}{2}\sin^2\theta)\,k^{\frac{2}{3}}K^{-\frac{2}{3}}, \tag{4·7}$$

$$\beta = \tfrac{1}{2}\sin^2\theta\,k^{\frac{2}{3}}K^{-\frac{2}{3}}, \tag{4·8}$$

we have

$$\frac{d^2L}{d\zeta^2}+(\alpha+\zeta)\,L+\beta M = 0, \tag{4·9}$$

$$\frac{d^2M}{d\zeta^2}+(\alpha-\zeta)\,M+\beta L = 0. \tag{4·10}$$

When $\beta = 0$, these equations become independent of each other, and it is easily shown, by the method of Laplace transforms, that the following contour integrals are solutions, provided the contour is suitably specified:

$$L(\zeta) = \int \exp\left[\tfrac{1}{3}t^3+(\alpha+\zeta)\,t\right]dt, \quad M(\zeta) = \int \exp\left[\tfrac{1}{3}t^3+(\alpha-\zeta)\,t\right]dt.$$

These lead to the well-known solutions in terms of the Airy integral and its associated function (Miller 1946; Jeffreys & Jeffreys 1946).

A similar approach is possible when β is not equal to zero, but it is found that the general solution cannot then be obtained in terms of a single contour integral, and that the sum of several such integrals must be taken.

Consider the integrals

$$I(\zeta) = \int S(t)\,e^{-\zeta t}\,dt, \tag{4·11}$$

$$J(\zeta) = \int T(t)\,e^{-\zeta t}\,dt, \tag{4·12}$$

where the contour of integration will be defined later. Denote by $R_L[I(\zeta), J(\zeta)]$, $R_M[I(\zeta), J(\zeta)]$ the remainders when $I(\zeta), J(\zeta)$ are substituted for $L(\zeta), M(\zeta)$ in the left-hand sides of (4·9) and (4·10). Then

$$R_L[I(\zeta), J(\zeta)] = \frac{d^2 I(\zeta)}{d\zeta^2} + (\alpha+\zeta)\, I(\zeta) + \beta J(\zeta)$$

$$= \int [(\alpha+t^2)\, S(t) + \beta T(t)]\, e^{-\zeta t}\, dt + \zeta \int S(t)\, e^{-\zeta t}\, dt.$$

If the second term is integrated by parts this becomes

$$R_L[I(\zeta), J(\zeta)] = \int \left[(\alpha+t^2)\, S(t) + \frac{dS(t)}{dt} + \beta T(t)\right] dt - [S(t)\, e^{-\zeta t}]^*$$

$$= -[S(t)\, e^{-\zeta t}],^* \tag{4·13}$$

provided $$(\alpha+t^2)\, S + \frac{dS}{dt} + \beta T = 0. \tag{4·14}$$

Similarly $$R_M[I(\zeta), J(\zeta)] = [T(t)\, e^{-\zeta t}],^* \tag{4·15}$$

provided $$(\alpha+t^2)\, T - \frac{dT}{dt} + \beta S = 0. \tag{4·16}$$

The normal procedure in applying a Laplace transformation to obtain a solution of a pair of differential equations, would now be to specify the contour so that $R_L[I(\zeta), J(\zeta)]$ and $R_M[I(\zeta), J(\zeta)]$ are both zero.† This is not possible in the present case, but it will be seen later that if we define a finite series of functions $I_1(\zeta)$, $I_2(\zeta)$, etc., by the relations

$$I_j(\zeta) = \int_{\Gamma_j} S_j(t)\, e^{-\zeta t}\, dt, \quad J_j(\zeta) = \int_{\Gamma_j} T_j(t)\, e^{-\zeta t}\, dt,$$

where $S_j(t)$ and $T_j(t)$ satisfy (4·14) and (4·16), the integrals being taken along different contours, Γ_j, then it is possible to choose constants a_1, a_2, etc., so that

$$\Sigma a_j R_L[I_j(\zeta), J_j(\zeta)] = 0, \tag{4·17}$$

$$\Sigma a_j R_M[I_j(\zeta), J_j(\zeta)] = 0. \tag{4·18}$$

The functions $$L(\zeta) = \Sigma a_j I_j(\zeta), \quad M(\zeta) = \Sigma a_j J_j(\zeta)$$

then constitute a solution of (4·9) and (4·10).

It will be convenient at this stage to make a change of notation. Write

$$T = u(t) \exp\left(\tfrac{1}{3}t^3 + \alpha t\right), \tag{4·19}$$

---

* Square brackets are used to denote the difference of the values of the enclosed function at the two ends of the contour.

† Cf. Watson (1944, p. 160).

then from (4·16)
$$S = \frac{1}{\beta}\frac{du(t)}{dt}\exp\left(\tfrac{1}{3}t^3 + \alpha t\right). \tag{4·20}$$

Thus
$$I(\zeta) = \beta^{-1}\int \frac{du(t)}{dt}\exp\left[\tfrac{1}{3}t^3 - (\zeta - \alpha)t\right]dt, \tag{4·21}$$

$$J(\zeta) = \int u(t)\exp\left[\tfrac{1}{3}t^3 - (\zeta - \alpha)t\right]dt. \tag{4·22}$$

By substituting for S and T in (4·14) it may be shown that the equation satisfied by $u(t)$ is
$$\frac{d^2u}{dt^2} + 2(\alpha + t^2)\frac{du}{dt} + \beta^2 u = 0. \tag{4·23}$$

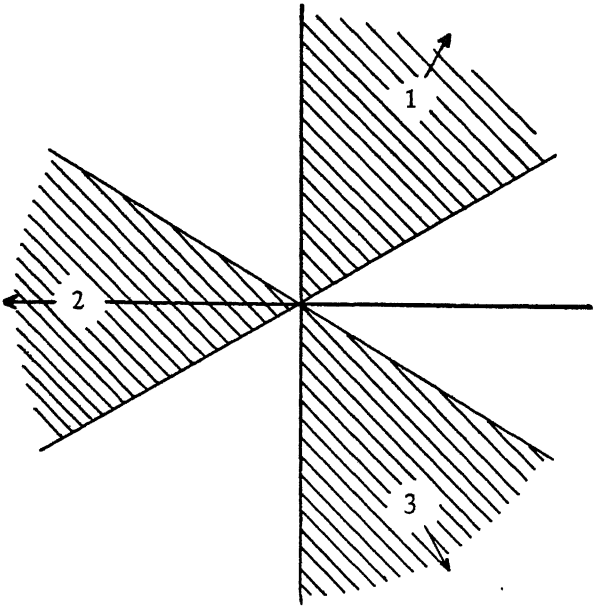


FIGURE 2

Equation (4·23) will be solved in a later section; for the time being it will be sufficient to say that it has an essential singularity at infinity, but that if attention is confined to one of the shaded sectors (numbered 1, 2, 3) in figure 2, it is possible to obtain a solution which is regular at all finite points of the sector, and which tends to a constant value at infinity. There are three solutions of this kind, $u_1(t)$, $u_2(t)$, $u_3(t)$, corresponding to the three sectors. It is not possible to obtain a solution which is regular and tends to zero at infinity in two specified sectors at the same time and such that u and du/dt are continuous at the origin. It is for this reason that $L(\zeta)$ and $M(\zeta)$ cannot be expressed in terms of a single integral along a contour starting at infinity in one sector and ending at infinity in another, as is possible in the case of Airy's integral.

Now define the following functions (cf. equations (4·21) and (4·22)):

$$I_n(\zeta) = \beta^{-1}\int \frac{du_n(t)}{dt} \exp\left[\tfrac{1}{3}t^3 - (\zeta-\alpha)\,t\right] dt, \tag{4·24}$$

$$J_n(\zeta) = \int u_n(t) \exp\left[\tfrac{1}{3}t^3 - (\zeta-\alpha)\,t\right] dt, \tag{4·25}$$

where n has the value of 1, 2 or 3. The contour of integration is any line, which, lying wholly in the sector n, starts at the origin, and ends at infinity.

If use is made of equations (4·13), (4·15), (4·19) and (4·20) it follows that

$$R_M[I_n(\zeta), J_n(\zeta)] = -u_n(0), \tag{4·26}$$

$$\begin{aligned} R_L[I_n(\zeta), J_n(\zeta)] &= \beta^{-1}\left[\frac{du_n(t)}{dt}\right]_{t=0} \\ &= \beta^{-1}u_n'(0). \end{aligned} \tag{4·27}$$

It will be noted that these remainders are independent of ζ.

By restricting attention to three sectors only of the complex plane, and by imposing the condition that $u_n(t)$ tends to a constant value at infinity, some loss of generality is incurred. This may be restored in the following manner. Note that equations (4·9) and (4·10) are interchanged if L and M are interchanged and $-\zeta$ is written for ζ. Define

$$I_n'(\zeta) = J_n(-\zeta), \tag{4·28}$$

$$J_n'(\zeta) = I_n(-\zeta). \tag{4·29}$$

It is easily shown that

$$R_M[I_n'(\zeta), J_n'(\zeta)] = \beta^{-1}u_n'(0), \tag{4·30}$$

$$R_L[I_n'(\zeta), J_n'(\zeta)] = -u_n(0). \tag{4·31}$$

It will be seen in § 4(*c*), from a consideration of the asymptotic forms, that the six pairs of functions

$$I_n(\zeta), \quad J_n(\zeta), \quad I_n'(\zeta), \quad J_n'(\zeta)$$

are linearly independent. Now take a linear combination of the functions, and write

$$L(\zeta) = \Sigma a_n I_n(\zeta) + \Sigma a_n' I_n'(\zeta), \tag{4·32}$$

$$M(\zeta) = \Sigma a_n J_n(\zeta) + \Sigma a_n' J_n'(\zeta), \tag{4·33}$$

where a_n, a_n' are independent of ζ. A solution of equations (4·9) and (4·10) follows provided

$$\Sigma a_n \beta^{-1} u_n'(0) - \Sigma a_n' u_n(0) = 0, \tag{4·34}$$

$$\Sigma a_n u_n(0) - \Sigma a_n' \beta^{-1} u_n'(0) = 0. \tag{4·35}$$

Moreover, since any four of the six quantities a_n, a_n' may be specified arbitrarily, it follows that the expressions (4·32) and (4·33) constitute the general solution of (4·9) and (4·10).

A formal solution of the problem in terms of a solution of the second order differential equation (4·23) for $u(t)$ has thus been obtained. The next section will be devoted to the study of this equation.

(*b*) *Solution of the equation for $u(t)$*

In all cases to which this analysis is applied β^2 is a small quantity. It is therefore natural to seek a solution of equation (4·23) in powers of β^2. Accordingly, take the trial solution

$$u_n(t) = 1+\beta^2 f_{n,1}(t)+\beta^4 f_{n,2}(t)+\dots\beta^{2r}f_{n,r}(t)+\dots, \tag{4·36}$$

where $f'_{n,1}(t), f'_{n,2}(t)$, etc., all tend to zero as $|t|$ tends to infinity in the sector n, so that $u_n(t)$ tends to a constant value. Substituting and equating coefficients of powers of β^2, we have

$$f''_{n,1}(t)+2(\alpha+t^2)f'_{n,1}(t)+1 = 0, \tag{4·37}$$

$$f''_{n,2}(t)+2(\alpha+t^2)f'_{n,2}(t)+f_{n,1}(t) = 0, \tag{4·38}$$

....

$$f''_{n,r}(t)+2(\alpha+t^2)f'_{n,r}(t)+f_{n,r-1}(t) = 0, \tag{4·39}$$

where the dashes denote differentiation with respect to the argument. These are all linear differential equations of the first order, and may be solved by means of an integrating factor. Hence

$$f'_{n,1}(t) = -\exp[-2(\tfrac{1}{3}t^3+\alpha t)]\int_{t_1}^{t}\exp[2(\tfrac{1}{3}s^3+\alpha s)]\,ds, \tag{4·40}$$

$$f'_{n,r}(t) = -\exp[-2(\tfrac{1}{3}t^3+\alpha t)]\int_{t_r}^{t}f_{n,r-1}(s)\exp[2(\tfrac{1}{3}s^3+\alpha s)]\,ds, \tag{4·41}$$

where the integrals are taken along contours in the s plane starting at the points t_1, t_r and ending at the point t.

As mentioned above it will be sufficient to confine attention to the domains of the t plane which are shaded in figure 2; it will be shown that if the point t_r is suitably chosen, $f_{n,r}(t)$ exists and tends to zero as $|t|$ tends to infinity.

$$f'_{n,1}(t) \sim e^{-\frac{2}{3}t^3}\int_{t_1}^{t}e^{\frac{2}{3}s^3}\,ds$$

$$\sim \frac{e^{-\frac{2}{3}t^3}}{(-18)^{\frac{1}{3}}}\left[\int_0^{-\frac{2}{3}t^3}\xi^{-\frac{2}{3}}e^{-\xi}\,d\xi-\int_0^{-\frac{2}{3}t_1^3}\xi^{-\frac{2}{3}}e^{-\xi}\,d\xi\right] \quad \text{when } |t| \text{ is large,}$$

$$\sim \frac{e^{-\frac{2}{3}t^3}}{(-18)^{\frac{1}{3}}}\left[\Gamma(\tfrac{1}{3})-(-\tfrac{2}{3})^{-\frac{2}{3}}t^{-2}e^{\frac{2}{3}t^3}-\int_0^{-\frac{2}{3}t_1^3}\xi^{-\frac{2}{3}}e^{-\xi}\,d\xi\right],$$

using the well-known asymptotic expansion for the incomplete Gamma function. Take t_1 to be the point at infinity in the direction indicated by the arrow in the particular shaded sector under discussion, i.e. $t_1 = \epsilon_n\infty$, where

$$\epsilon_n = \exp[\tfrac{1}{3}(2n-1)\pi i],$$

so that

$$\int_0^{-\frac{2}{3}t_1^3}\xi^{-\frac{2}{3}}e^{-\xi}\,d\xi = \Gamma(\tfrac{1}{3}),$$

and hence

$$f'_{n,1}(t) \sim -\tfrac{1}{2}t^{-2}. \tag{4·42}$$

It follows that as $|t|$ tends to infinity in the shaded sector under consideration

$$f_{n,1}(t) \sim A_1 + \tfrac{1}{2}t^{-1},$$

where A_1 is a constant, depending on the choice of the lower limit of integration. It is convenient to take this to be the point $t = 0$, so that

$$f_{n,1}(t) = -\int_0^t \exp\left[-2(\tfrac{1}{3}r^3 + \alpha r)\right] dr \int_{\epsilon_n\infty}^{r} \exp\left[2(\tfrac{1}{3}s^3 + \alpha s)\right] ds. \tag{4·43}$$

There is no loss of generality in taking $t = 0$ for the lower limit of integration, since it may be shown that the effect of taking any other value would be merely to multiply the value of $u_n(t)$ given by (4·36) by a constant.

Return now to (4·41), take $t_r = \epsilon_n \infty$, and integrate $f'_{n,r}(t)$ along a contour starting at the origin.

$$f_{n,r}(t) = -\int_0^t \exp\left[-2(\tfrac{1}{3}r^3 + \alpha r)\right] dr \int_{\epsilon_n\infty}^{r} f_{n,r-1}(s) \exp\left[2(\tfrac{1}{3}s^3 + \alpha s)\right] ds. \tag{4·44}$$

It may be shown by induction that $f_{n,r}(t)$ exists and tends to the form $A_r + \tfrac{1}{2}t^{-1}$ as $|t|$ tends to infinity.

In evaluating the integrals in equations (4·24), (4·25) values of $f_{n,r}(t)$ will only be required for points lying on straight lines radiating from the origin at angles $\tfrac{1}{3}\pi$, π and $\tfrac{5}{3}\pi$ to the real axis, i.e. in the direction of the arrows in figure 2, and it is therefore convenient to write

$$r = \rho\epsilon_n e^{i\theta}, \quad s = \sigma\epsilon_n e^{i\theta}, \quad t = \tau\epsilon_n e^{i\theta}, \tag{4·45}$$

where $|\theta| < \tfrac{1}{6}\pi$ and ρ, σ, τ are real and positive. Ultimately the values of $f_{n,r}(t)$ for which $\theta = 0$ will be required.

It will now be proved that the series (4·36) is uniformly convergent with respect to t in a domain of the complex plane containing the line $\theta = 0$, provided that β^2 is less than a certain quantity.

$$f_{n,1}(\tau\epsilon_n e^{i\theta}) = -\epsilon_n^2 e^{2i\theta} \int_0^\tau \exp\left[\tfrac{2}{3}\rho^3 e^{3i\theta} - 2\alpha\rho\epsilon_n e^{i\theta}\right] d\rho \int_\rho^\infty \exp\left[-\tfrac{2}{3}\sigma^3 e^{3i\theta} + 2\alpha\sigma\epsilon_n e^{i\theta}\right] d\sigma.$$

Taking the moduli of the integrands, and replacing the upper limit of the integration in ρ by infinity, we have

$$\begin{aligned} |f_{n,1}(\tau\epsilon_n e^{i\theta})| &\leqslant \int_0^\infty \exp\left[\tfrac{2}{3}\rho^3 \cos 3\theta - 2\alpha\rho \cos(\tfrac{1}{3}\overline{2n-1}\pi + \theta)\right] d\rho \\ &\qquad \int_0^\infty \exp\left[-\tfrac{2}{3}\sigma^3 \cos 3\theta + 2\alpha\sigma \cos(\tfrac{1}{3}\overline{2n-1}\pi + \theta)\right] d\sigma \\ &= \sec^{\frac{2}{3}} 3\theta \int_0^\infty \exp\left[\tfrac{2}{3}\rho^3 - 2\alpha\rho \sec^{\frac{1}{3}} 3\theta \cos(\tfrac{1}{3}\overline{2n-1}\pi + \theta)\right] d\rho \\ &\qquad \int_\rho^\infty \exp\left[-\tfrac{2}{3}\sigma^3 + 2\alpha\sigma \sec^{\frac{1}{3}} 3\theta \cos(\tfrac{1}{3}\overline{2n-1}\pi + \theta)\right] d\sigma \\ &= K(\alpha, \theta). \end{aligned}$$

That $K(\alpha, \theta)$ exists may be proved by an argument similar to that used to prove the existence of $f_{n,1}(t)$.

By replacing $f_{n,1}(s)$ in the integral for $f_{n,2}(s)$ by $K(\alpha, \theta)$ it is easy to see that

$$|f_{n,2}(t)| < [K(\alpha, \theta)]^2,$$

and continuing the process that

$$|f_{n,r}(t)| < [K(\alpha, \theta)]^r.$$

It now follows that the series (4·36) converges uniformly with respect to t in the domain defined by $-\theta_0 < \theta < \theta_0$ provided that $\beta^2 < 1/\bar{K}(\alpha, \theta_0)$, where $\bar{K}(\alpha, \theta_0) =$ upper bound of $K(\alpha, \theta)$ in the domain. $\bar{K}(\alpha, \theta_0)$ is a continuous function of θ_0, and if

$$\beta^2 < A < 1/K(\alpha, 0),$$

where A is a constant, there exists a domain, containing the line $\theta = 0$, in every bounded subdomain of which the series (4·36) is convergent uniformly with respect to t. Within this domain, differentiation of the series (4·36) term by term is legitimate, and its sum therefore represents a solution of equation (4·23).

Thus by taking n in equation (4·44) equal to 1, 2, 3 in turn and substituting for $f_{n,r}(t)$ in (4·36), expressions are obtained for the quantities denoted in § 4 (a) by $u_1(t)$, $u_2(t)$, $u_3(t)$.

These may now be substituted in equations (4·24), (4·25) to give $I_n(\zeta)$, $J_n(\zeta)$ from which solutions of the differential equations (4·9) and (4·10), governing propagation in the reflecting region, may be constructed in the manner explained at the end of § 4 (a).

(c) *Asymptotic forms*

The following asymptotic formulae valid when ζ is large may now be established:

$$J_1(\zeta) \sim \frac{u_1(0)}{\zeta} + \frac{u_1'(0) + \alpha u_1(0)}{\zeta^2} + 0\left(\frac{1}{\zeta^3}\right), \tag{4·46}$$

$$I_1(\zeta) \sim \beta^{-1}\left[\frac{u_1'(0)}{\zeta} + \frac{u_1''(0) + \alpha u_1'(0)}{\zeta^2} + 0\left(\frac{1}{\zeta^3}\right)\right], \tag{4·47}$$

$$J_2(\zeta) \sim u_2(\infty)\sqrt{\pi}(\zeta - \alpha)^{-\frac{1}{4}} \exp\left[\tfrac{2}{3}(\zeta - \alpha)^{\frac{3}{2}}\right], \tag{4·48}$$

$$I_2(\zeta) \sim \beta^{-1} C_2 \sqrt{\pi}(\zeta - \alpha)^{-\frac{5}{4}} \exp\left[\tfrac{2}{3}(\zeta - \alpha)^{\frac{3}{2}}\right], \tag{4·49}$$

$$J_3(\zeta) \sim \frac{u_3(0)}{\zeta} + \frac{u_3'(0) + \alpha u_3(0)}{\zeta^2} + 0\left(\frac{1}{\zeta^3}\right), \tag{4·50}$$

$$I_3(\zeta) \sim \beta^{-1}\left[\frac{u_3'(0)}{\zeta} + \frac{u_3''(0) + \alpha u_3'(0)}{\zeta^2} + 0\left(\frac{1}{\zeta^3}\right)\right], \tag{4·51}$$

$$J_1'(\zeta) \sim \beta^{-1} C_1 \sqrt{\pi}(\zeta + \alpha)^{-\frac{5}{4}} \exp\left[\tfrac{2}{3}(\zeta + \alpha)^{\frac{3}{2}} i + \tfrac{1}{4}\pi i\right], \tag{4·52}$$

$$I_1'(\zeta) \sim u_1(\infty)\sqrt{\pi}(\zeta+\alpha)^{-\frac{1}{4}}\exp\left[\tfrac{2}{3}(\zeta+\alpha)^{\frac{3}{2}}i+\tfrac{1}{4}\pi i\right], \tag{4·53}$$

$$J_2'(\zeta) \sim \beta^{-1}\left[-\frac{u_2'(0)}{\zeta}+\frac{u_2''(0)+\alpha u_2'(0)}{\zeta^2}+0\left(\frac{1}{\zeta^3}\right)\right], \tag{4·54}$$

$$I_2'(\zeta) \sim -\frac{u_2(0)}{\zeta}+\frac{u_2'(0)+\alpha u_2(0)}{\zeta^2}+0\left(\frac{1}{\zeta^3}\right), \tag{4·55}$$

$$J_3'(\zeta) \sim \beta^{-1}C_3\sqrt{\pi}(\zeta+\alpha)^{-\frac{1}{4}}\exp\left[-\tfrac{2}{3}(\zeta+\alpha)^{\frac{3}{2}}i-\tfrac{1}{4}\pi i\right], \tag{4·56}$$

$$I_3'(\zeta) \sim u_3(\infty)\sqrt{\pi}(\zeta+\alpha)^{-\frac{1}{4}}\exp\left[-\tfrac{2}{3}(\zeta+\alpha)^{\frac{3}{2}}i-\tfrac{1}{4}\pi i\right]. \tag{4·57}$$

In these formulae, $u_n(\infty)$ denotes the limiting value of $u_n(t)$ as $|t|$ tends to infinity, and C_n is a constant defined by

$$u_n'(t) \sim C_n/t^2$$

when $|t|$ is large.

The formula for $J_1(\zeta)$ may be proved as follows. Write

$$g_1(t) = u_1(t)\exp\left(\tfrac{1}{3}t^3+\alpha t\right);$$

then, by repeated integration by parts,

$$\begin{aligned} J_1(\zeta) &= \int_0^{\epsilon_1\infty} e^{-\zeta t}g_1(t)\,dt \\ &= \frac{g_1(0)}{\zeta}+\frac{g_1'(0)}{\zeta^2}+\frac{g_1''(0)}{\zeta^3}+\frac{1}{\zeta^r}\left[g^{(r-1)}(0)+\int_0^{\epsilon_1\infty} e^{-\zeta t}g^{(r)}(t)\,dt\right]. \end{aligned}$$

Since the term in $1/\zeta^r$ tends to zero as ζ tends to infinity, this is an asymptotic series for $J_1(\zeta)$. Expressed in terms of $u_1(t)$, the first two terms reduce to form (4·46). A similar argument may be used to establish the validity of the expansions (4·47), (4·50), (4·51), (4·54) and (4·55).

The remaining expressions may be obtained by the method of steepest descents. Consider, for example,

$$J_2(\zeta) = \int_0^{-\infty} u_2(t)\exp\left[\tfrac{1}{3}t^3+(\alpha-\zeta)t\right]dt.$$

The exponential function has saddle points at $t = \pm(\zeta-\alpha)^{\frac{1}{2}}$ of which one—corresponding to the lower sign—lies in the sector 2; if the contour of integration is taken to pass through this point along a line of steepest slope, the expression given in (4·48) is easily obtained. The same method may be applied to obtain (4·49).

There is a slight complication in applying the method of steepest descents to the other functions, for example, $J_1'(\zeta) = I_1(-\zeta)$ because the saddle points of the exponential function occur at the points $t = \pm(\zeta+\alpha)^{\frac{1}{2}}i$, which lie on the boundaries of the sectors within which $u(t)$ is finite for large t. It may, however, be shown that when ζ is large, the important contribution to each integral comes from those portions of the contours which are remote from the origin, and that the leading terms in the asymptotic expansions may be obtained by putting in the limiting forms for $u_n(t)$ and $u'_n(t)$, i.e. $u_n(\infty)$ and C_n/t^2 and evaluating the resulting integrals by the method of steepest descents in the usual manner.

A study of the asymptotic forms given in this section shows that, as already stated in § 4 (*a*), the six pairs of functions $I_n(\zeta), J_n(\zeta), I'_n(\zeta), J'_n(\zeta)$ are linearly independent.

5. APPLICATION TO THE IONOSPHERE

A solution of (4·9) and (4·10) representing the most general wave which can be propagated in the reflecting region has now been obtained. The next step in applying the analysis to the ionosphere is to impose on this solution two conditions, (*a*) that it should be finite at infinity, and (*b*), that it should be continuous, at the lower boundary of the reflecting region, with expressions representing the waves (upgoing and downgoing) propagated in the region immediately below. If the effect of the transition region is neglected, the latter is equivalent to free space, and the analysis follows the lines of that in Wilkes (1940, p. 149), but generalized to allow for oblique incidence.

In order to carry out the procedure outlined above, it is necessary to evaluate the functions $I_n(\zeta), J_n(\zeta)$ and their derivatives at the lower boundary of the reflecting region, where ζ is zero or has some given small value. If β is sufficiently small for its second and higher power to be neglected, only two terms of the series for $u_n(t)$ are needed, and it is found that $I_n(\zeta), J_n(\zeta)$ may be expressed by means of power series in a form suitable for computation. This approximation is sufficiently accurate for the discussion of the experimental results on the propagation of 16 kyc./sec. waves referred to in the Introduction, since it will be found on substitution in (4·8) that even when θ is as great as 60°, β is only 0·13, while for smaller values of θ it is even smaller. On higher frequencies the approximation obtained by neglecting β^2 will correspond to reflexion at small angles of incidence.

The calculations will not be carried through in detail in the present paper, but it is found that the four linearly independent expressions for $L(\zeta)$ and $M(\zeta)$ obtained from (4·32) and (4·33) by assigning values to the six constants a_n, a'_n, subject to the two conditions (4·34) and (4·35), are equivalent to the solution for $\beta = 0$ plus a correcting term of order β. It follows that the error in calculating the wave reflected from the medium on the basis of a solution for vertical incidence is of the order β. One is therefore justified in using the results of the analysis of Wilkes (1940) to give a first approximation in the case of oblique reflexion at small angles of incidence.

I would like to express my thanks to Dr F. Smithies for several helpful discussions and to Mr J. A. Ratcliffe for advice as to the final form the paper should take. I am also indebted to Professor D. R. Hartree, F.R.S., Dr H. G. Booker and Mr G. Millington for reading the manuscript and commenting on it, and to Dr and Mrs Jeffreys for kindly allowing me to see in proof certain sections of their forthcoming book.

REFERENCES

Best, Ratcliffe & Wilkes 1936 *Proc. Roy. Soc.* A, **156**, 614.
Booker 1934 *Proc. Roy. Soc.* A, **147**, 352.
Booker 1935 *Proc. Roy. Soc.* A, **150**, 267.
Booker 1936 *Proc. Roy. Soc.* A, **155**, 235.
Booker 1938 *Phil. Trans.* A, **237**, 411.
Budden, Ratcliffe & Wilkes 1939 *Proc. Roy. Soc.* A, **171**, 188.
Chapman 1931*a* *Proc. Phys. Soc.* **43**, 26.
Chapman 1931*b* *Proc. Phys. Soc.* **43**, 483.
Chapman 1939 *Proc. Phys. Soc.* **51**, 93.
Försterling & Lassen 1933 *Ann. Phys., Lpz.*, **18**, 26.
Hartree 1931 *Proc. Roy. Soc.* A, **131**, 428.
Havelock 1914 *Camb. Math. Tract.* no. 17.
Jeffreys & Jeffreys 1946 *Methods of Mathematical Physics*. Camb. Univ. Press.
Miller, J. C. P. 1946 *British Association Mathematical Tables*, Part-volume B, prepared by J. C. P. Miller.
Taylor 1934 *Proc. Phys. Soc.* **46**, 408.
Watson 1944 *Theory of Bessel Functions*, 2nd ed.. Camb. Univ. Press.
Wilkes 1940 *Proc. Roy. Soc.* A, **175**, 143.

The fine structure of the spectral lines of H_2: measurements with a reflexion echelon

By E. W. Foster and Sir Owen Richardson, F.R.S.

(*Received* 5 *December* 1944)

[Plate 4]

The fine structure of the lines of the H_2 spectrum was discovered by Richardson & Williams in 1931, and a short account of their observations was published in a letter to *Nature* at the time. Since then no further publication on the subject has appeared. The present paper represents the results of experiments and observations made during the last six years with improved apparatus, including the Royal Society's 40 plate reflexion echelon. About 250 lines at the red end of the visible spectrum have been examined and their fine structures photographed and measured. These include most of the important lines of the triplet systems $3p\,{}^3\Pi \rightarrow 2s\,{}^3\Sigma$ and $3d\,{}^3\Sigma\Pi\Delta \rightarrow 2p\,{}^3\Pi$, as well as important lines of the system $3s\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$ and $3\,Y \rightarrow 2s\,{}^3\Sigma$, together with some lines of the triplet system $3p\,{}^3\Sigma \rightarrow 2s\,{}^3\Sigma$. The singlets are represented by about 20 lines coming from the states $3d\,{}^1\Sigma$, $3d\,{}^1\Pi_d$, $3s\,{}^1\Sigma$, $2s\,{}^1\Sigma$, $3\,{}^1K$, ${}^1M$, and ${}^1N$ and ending in $2p\,{}^1\Sigma$ or belonging to the system $3p\,{}^1\Pi \rightarrow 2s\,{}^1\Sigma$. There is also a small proportion of unclassified lines.

The present paper is limited to a descriptive account of the observations and the analysis and tabulation of the data obtained. The analysis of these data has already been carried out and the results will be published separately.

The plate which reproduces a photograph of the crossed spectra of the stronger lines dealt with is the first published picture showing the fine structure under consideration.

1. Introduction

This fine structure was discovered by Richardson & Williams in 1931 and the only published description of it is in a letter to *Nature* (Richardson & Williams 1931) which is brief and provisional. The observations were made with a system consisting of a large Hilger quartz spectrograph crossed with a reflexion echelon. Those now to be described were made by the same method using a larger echelon (40 plates) and other improvements. The apparatus was, in fact, the same as that used by Drinkwater, Richardson & Williams (1940) to obtain the results published in the paper on 'Determinations of the Rydberg constants, e/m, and the fine structures of H_α and D_α by means of a reflexion echelon' apart from modifications adopted to meet the changed conditions of excitation and character of the lines. Altogether 244 lines measured by Gale, Monk & Lee (1928) have been photographed and measured and are here described and analysed. Actually the number of lines is considerably more than 244, as many of these are blends. This often increased the difficulty of interpretation. A photograph of the fine-structure patterns of the strongest of them is reproduced in plate 4. They lie between the wave numbers 17741·92 and 15443·06 (λ 5634·807(8) to 6473·615(2)).

The photographing of the spectra was all done in 1938–39 and was brought to a sudden end by the events of September 1939, which included the evacuation of

King's College, London. This accounts for the limited range of the spectrum represented. However, it is much the most interesting and important part, having regard to the resolution which can be brought to bear in practice on this spectrum at the present time. There were other regions which we should have looked into had there been time.

The work subsequent to September 1939 was carried on mainly at Chandos Lodge, Alton, Hampshire. This included all measurements of the plates and their reduction, the enlarging and photometry and the analysis and interpretation of the resulting data.

The scope of this paper is confined to a description of the experimental results. The delay in publication is due to the frustrations and interruptions which are inevitable under conditions of total war. The analysis of these results will be dealt with in subsequent papers.

2. Spectrographic equipment and methods

The spectrographic equipment consisted of a reflexion echelon (40 plates, step width 6·88 mm.) and a Hilger (E 1 type) Littrow spectrograph fitted with a glass optical train. The two instruments were crossed so that the high dispersion of the echelon was in a vertical plane perpendicular to the auxiliary coarse dispersion of the Littrow instrument. Since the particular dispersing apparatus used by us has already been described in detail (Williams 1933; Williams & Middleton 1939) we shall not here do more than briefly refer to it.

A mirror is mounted at each side of the echelon almost parallel to the reflecting faces of the steps. The normals to these mirrors are slightly inclined on opposite sides of the normal to the reflecting faces of the echelon so that the images of the echelon collimator slit in the mirrors provide a series of reference marks (M_1, M_2, plate 4) which are used for plate measurement. These marks also determine the change in the magnification of the optical system with wave-length.

The echelon is mounted in a vacuum chamber and molecular lines were measured directly (§ 6) in terms of the vacuum wave number of the standard Cd red line λ_{air} 6438·4696 A or $\lambda_{\text{vac.}}$ 6440·2491 A. Additional measurements were made with the vacuum chamber filled with air at a pressure of a few centimetres of mercury. The inter-order dispersion was obtained from measurements of the separation of very sharp argon lines appearing in approximately the double-order position.

On each plate it was usual to record the standard Cd line several times with different exposures both before and after an exposure of the molecular spectrum. The argon spectrum also was usually exposed after the molecular spectrum. Unless there was good reason to do otherwise a plate was discarded if the Cd exposures before and after were not in agreement. This happened in only one case (when it is believed that some mechanical shock had disturbed the relative orientation of the echelon and its mirrors) after precautions to keep the laboratory temperature constant (± 0·1° C) had been taken.

The molecular spectrum was photographed under high dispersion between about 5000 and 6500 A, using Ilford 'Hypersensitive Panchromatic' plates. The faintest structures recorded required an exposure of about 30 hr.

The spectrum was measured between $\nu = 17859{\cdot}76$ and $\nu = 15443{\cdot}06$. The method of computing the wave numbers of the rather broad molecular lines is dealt with in § 6.

3. EXCITATION OF THE MOLECULAR HYDROGEN SPECTRUM

The H_2 spectrum was observed 'end on' through a window on one of the 30 cm. long vertical limbs of a U-shaped bend in the central part of a pyrex discharge tube, of overall length 180 cm. and uniform internal diameter about 1 cm. The ends of the tube widened to contain hollow cylindrical thin sheet-nickel electrodes, 12 cm. long, 2·5 cm. diameter.

Arrangements were provided for exhausting the tube by two-stage mercury diffusion pumps, one at each end of the tube. Liquid air-cooled traps were provided to prevent mercury vapour from entering the discharge. At one end of the discharge tube pure hydrogen could be admitted from a reservoir by diffusion through an electrically heated palladium tube. A small McLeod-type gauge was joined close to the pump connexion at the other end. A liquid air-cooled charcoal trap could also be connected to the discharge tube to counteract air leakage; it had the disadvantage of slowly absorbing hydrogen, but it was connected to the discharge tube during a long exposure as a precaution.

The discharge tube and electrodes were brought to a very clean state by the prolonged passage of a discharge in hydrogen with repeated pumping and flushing out with pure gas. In the early stages the discharge had the usual pinkish tinge, sometimes with striations, and H_α was very strong compared with any molecular lines in the visible spectrum. When the last traces of impurity were gone the positive column became white everywhere, and in this condition, which was always used when photographing the molecular spectrum, the strongest yellow and red molecular lines were comparable in visual intensity with H_α. Subsequently no impurities of any description were ever observed in the hydrogen discharge unless purposely admitted.

For photographing the molecular spectrum the tube contained hydrogen at a pressure of about 0·15 mm. Hg, which was excited to the white stage by a direct current of about 20 mA. The U-shaped bend in the tube was conveniently cooled by immersion in a deep Dewar flask containing liquid oxygen. This sharpened the molecular lines very considerably and enabled the resolution of many hitherto unknown structures.

4. EXCITATION OF THE ARGON SPECTRUM

Very sharp argon lines were easily and conveniently excited in the hydrogen tube described in § 3. To this was attached a reservoir of spectroscopically pure argon, and a few flushings with this gas while passing an alternating current discharge, at about 100 mA, sufficed to remove residual hydrogen almost completely (H_α appeared very faintly). The converse process more rigorously applied was used to remove argon before an H_2 exposure. When photographing the argon spectrum the discharge tube bend was cooled in liquid oxygen.

5. EXCITATION OF THE STANDARD CADMIUM RED LINE

The Cd red line was excited in a water-cooled hollow cathode tube very similar to that described by Schüler (1931). The iron cathode, about 3 cm. deep, 1 cm. bore, contained a few pieces of metallic cadmium. It was found that the red line could be strongly excited in such a tube (exposure time about 2 min.) at lower current (about 80 mA) using pure argon as the carrier gas instead of the more usual helium (about 200 mA), presumably due to the greater sputtering of the cadmium in argon. This resulted in such an obviously sharper contour for the red line that argon was adopted in spite of the more elaborate gas purification system required.

Pure argon, at a pressure of about 0·05 mm. Hg, was circulated continuously through the Cd tube. The purification was effected by including in the circulator system a second (*hot*-cathode) Schüler tube containing metallic calcium. This tube passed a current of several amperes and effectively absorbed the inevitable residue of oxygen and nitrogen. Residual hydrogen was removed in the usual manner by passing the gas through a quartz tube containing red-hot copper oxide. Water vapour and remaining impurities were condensed in a liquid oxygen-cooled trap. The continuous circulation was effected by a two-stage mercury diffusion pump.

The wave number of the Cd line as excited by this source was not compared with the wave number from a standard source. Williams & Gogate (1938), however, have compared the wave number from a Cd-He Schüler tube indirectly with a standard Cd source and have found no appreciable difference. We have assumed that the substitution of argon for helium carrier gas will make no difference to the wave number of the Cd line.

6. MEASUREMENT AND REDUCTION OF PLATES

As will be seen from plate 4 (the dispersion is 0·726 cm^{-1} per order) the molecular lines are quite broad and, for reasons discussed below, it was not possible, in general, to reduce measurements directly by the simple method devised by Williams (1931) for sharp lines and thereby obtain results of the desired accuracy.

The equation of the echelon intensity envelope may be written

$$I = \frac{\sin^2 X}{X^2}, \quad \text{where} \quad X = \frac{\pi s(\alpha+\beta)}{\lambda},$$

s being the echelon step width, λ the wave-length and α and β the (very small) angles of incidence and reflexion, for the reflecting surface of the step. When $X = 0$ a line is said to be in single-order position; in which case, if the natural intensity distribution in the line is symmetrical, it will remain symmetrical in the diffraction pattern. For all other values of X between $\pm\pi$ the original symmetry will be distorted by the envelope effect.

It appears that the position of the 'optical centre of gravity' of a distorted component in the echelon pattern can be measured to a good approximation by the method described in the following paragraph. It is not obvious why this should be so, but we adopted the method in spite of its empirical nature since with the time and apparatus available it would have been impossible to microphotometer more than a small fraction of the lines actually measured, and owing to the considerable blending in the coarse dispersion it is doubtful whether the accuracy of the visual measures could have been improved much in very many cases.

Measurements were made with a specially constructed Hilger measuring microscope (Williams & Middleton 1939), the unusual feature of which was a travelling stage. On any one plate the position of a component was measured with respect to the reference marks M_1 and M_2 associated with the line in question. As a rule, twelve settings were made on the visually estimated centre of intensity of the component and six on each of the (considerably sharper) associated reference marks. Six of the settings on the component and three of those on the reference marks were made traversing the plate in the direction of increasing order of interference, and corresponding settings were made while traversing in the opposite direction. To change from the first measuring position to the second, the plate was rotated through 180° about a normal axis so that the microscope stage required moving in the same direction, whether for measurements in the direction of increasing or decreasing order of interference. The mean of the two sets of measures was taken. When a component appeared in more than one order all orders were measured, but it seldom happened that more than two orders were visible. The weakest lines were measured on only one plate but the majority have been measured on five or six plates.

The problem remaining was to relate the position thus obtained for the 'optical centre of gravity' of the component with the position which the component would assume if its width is imagined to become vanishingly small, i.e. with the limiting position of the line, as the temperature of the source tends to zero. When this is solved the calculation of wave number may be concluded by Williams's method referred to above.

The procedure adopted was as follows:

(i) The characteristic linear inter-order dispersion was obtained by measurement of several sharp argon lines appearing in almost symmetrical double-order position,* $X = \pm\frac{1}{2}\pi$ (when the intensities in the two orders are equal). These dispersions were reduced to the value for the Cd red line, and it was assumed that this was the dispersion for infinitely sharp lines. Similar measurements were made on a number of molecular lines also appearing in (or very nearly in) the double-order position (the equality of intensity was checked microphotometrically), and the apparent linear dispersion at λ Cd red was computed for these lines also. The value was less than the value for the sharp argon lines by about 0·02 order. One-half the difference between the argon and hydrogen order separations was taken as the displacement of a typical hydrogen line towards the centre of the intensity envelope at $X = \pm \frac{1}{2}\pi$. We refer later to this quantity as the 'half-order defect'.

---

* Cases where this is approximately realized for lines of the H_2 spectrum can be seen on plate 4; for the following lines of $3p\,{}^3\Pi \rightarrow 2s\,{}^3\Sigma$: 15552·37 (10), $4 \rightarrow 4$, $Q\,1$, 15859·86 (5), $3 \rightarrow 3$, $R\,0$, 16121·52 (8), $2 \rightarrow 2$, $R\,0$, 16314·99 (4), $0 \rightarrow 0$, $P\,4$, 16440·72 (10), $1 \rightarrow 1$, $R\,1$, 16611·43 (10), $0 \rightarrow 0$, $Q\,1$; for the $2 \rightarrow 0$, $R\,1$ line of $3p\,{}^3\Sigma \rightarrow 2s\,{}^3\Sigma$: 15669·39 (4); for the high-frequency component of the following lines ending on $2p\,{}^3\Pi$ (the low-frequency components which lie to the left are invisible in the left-hand orders owing to the envelope effect) 16458·03 (5), $3d\,{}^3\Pi_d \rightarrow 2p\,{}^3\Pi$, $3 \rightarrow 3$, $R\,1$, 16654·24 (9), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $0 \rightarrow 0$, $Q\,3$ and 16692·00 (5), $3d\,{}^3\Pi_d \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $P\,3$.

(ii) The position of the intensity maximum relative to the midpoint of the mirror marks was found approximately for any plate by determination of the distance between the position of the midpoint of the mirror marks and the midpoint of the orders of a line in double-order position. A line was considered to be in double-order position when microphotometer measures showed that the intensities in the two orders were equal. Having found the approximate position of the intensity maximum relative to the midpoint of the mirror marks it was a simple matter to find an approximate value of the fractional part f_I of every component referred to the intensity maximum, from the fractional part f_m derived directly from the measurements.

(iii) It was assumed that for a line having a fractional part f_I the displacement D towards the centre of the intensity curve would be proportional to the slope of that curve, i.e.

$$D = -K\frac{dI}{dX} = \frac{2K\sin X}{X^3}(X\cos X - \sin X)$$

at the point where the line occurred. D was plotted against X with the value of K adjusted so that the half-order defect was equal to D when $X = \frac{1}{2}\pi$.

(iv) The defect in the fractional part for a line in any position could then be found. This was added to the observed fractional part f_m giving the position which the line would have occupied had it been infinitely sharp. The remainder of the calculation was carried out by the method of Williams.

Results for the same line in different orders and in different positions (position varied by changing pressure in the echelon chamber) showed good agreement, so that for practical purposes the assumptions made were justified.

7. ON THE POSSIBILITY OF SYSTEMATIC ERRORS IN THE RESULTS FROM MEASUREMENTS ON PARTICULAR PLATES AND COMPARISON OF MEAN RESULTS WITH GALE, MONK AND LEE'S DATA

Gale *et al.* (1928) interferometer measures on eleven single, or substantially single, lines are compared with our individual results from particular plates and with the means of our results in table 1. The eleven lines are all members of the system $3p\,^3\Pi \rightarrow 2s\,^3\Sigma$. There is little doubt that these lines have some structure, but for practical purposes they are observationally simple with the resolution that can be brought to bear on them at present.

Wave numbers in italic type have been obtained from the measurement of one order of a line only; in ordinary type they have all been obtained on measurements on two orders.

The 'weighted mean of all our results = ν_F' may include results other than those given in the table. For comparison of results on individual plates, only results from measurements on the same number of orders for a given line on any one plate have been included.

TABLE 1

plate number →	40	83	84	85	86	weighted mean of all our results	
G.M.L.	our result (decimal part only)						defect
ν cm.$^{-1}$ = ν_G	(G.M.L. result) − (our result) = 'defect'					= ν_F	$\nu_G - \nu_F$
15552·37* (10)	·327	·334	·328	·320	·327	·327	·043
4 → 4, *Q* 1	·043	·036	·042	·050	·043	? ± ·008	
15621·96* (10)	·962	·963	·959	·958	·958	·960	·000
3 → 3, *P* 3	− ·002	− ·003	·001	·002	·002	± ·005	
15767·10* (7)	*·062*	*·071*	*·067*	*·070*	*·068*	*·067*	·033
3 → 3, *Q* 3	*·038*	*·029*	*·033*	*·030*	*·032*	± *·006*	
15787·23* (5)	*·218*	*·216*	*·213*	*·219*	*·206*	*·212*	·018
3 → 3, *Q* 2 and ?‡	*·012*	*·014*	*·017*	*·011*	*·024*	± *·008*	
15800·76* (10)	·739	·748	·740	·744	·746	·743	·017
3 → 3, *Q* 1 and ?‡	·021	·012	·020	·016	·014	± ·008	
15870·10* (10)	·066	·067	·068	·057	·058	·063	·037
2 → 2, *P* 3	·034	·033	·032	·043	·042	± ·006	
16046·28* (7)	*·249*	*·251*	*·253*	*·249*	*·258*	*·252*	·038
2 → 2, *Q* 2	*·031*	*·029*	*·027*	*·031*	*·022*	± *·006*	
16121·52* (8)	·490	·498	·494	·491	·486	·492	·028
2 → 2, *R* 0	·030	·022	·026	.029	·034	± ·004	
16168·95* (10)	*·915*	*·916*	*·914*	*·912*	*·918*	*·915*	·035
2 → 2, *R* 1	*·035*	*·034*	*·036*	*·038*	*·032*	± *·004*	
16330·59* (10)	*·567*	*·566*	*·568*	*·569*	*·571*	*·576*	·014
1 → 1, *Q* 1	*·023*	*·024*	*·022*	*·021*	*·019*	± *·006*	
16611·43* (10)	·405	·401	·400	·396	·395	·399	·031
0 → 0, *Q* 1	·025	·029	·030	·034	·035	± ·003	
mean defects	·026	·023	·026	·028	·028	—	·027

‡ See Table 3 for possible weak blends.

The wave number of the line 15,552·37 may be affected by the nearby 15549·47 (3) $3p\,{}^3\Pi \rightarrow 2s\,{}^3\Sigma$, 3 → 3, *P* 4 which does not appear on our plates.

From table 1, and an expanded table of the same form which includes lines not measured with the interferometer by Gale *et al.*, it is concluded that our plates give self-consistent results and that there is no systematic error relative to the mean result from all others, in any one of them $> c.\ 0{\cdot}002$ cm.$^{-1}$.

8. ON THE POSSIBILITY OF OUR MEAN RESULTS BEING SUBJECT TO A SYSTEMATIC ERROR

The H_2 lines were not measured directly in terms of the cadmium red line as an internal standard in the discharge tube, but one plate (no. 39) was obtained with a discharge atmosphere consisting of hydrogen and argon, and a few hydrogen lines given in table 2 were measured in terms of an accurately known argon line ($\nu = 15580{\cdot}984_0 \pm 0{\cdot}000_5$ cm.$^{-1}$) which had been measured previously against the cadmium line as internal standard (plate no. 79). The results should, for practical purposes, be as significant as if the H_2 lines had been measured against cadmium as an internal standard in a hydrogen discharge.

TABLE 2

ν_G	15552·37*	15767·10†	16126·19†	16440·72†	16611·43†
ν_F‡	·327	·067	·149	·693	·399
ν_{39}	·341	·065§	·144§	·690	·404
$\nu_G - \nu_F$	·043	·033	·041	·027	·031
$\nu_G - \nu_{39}$	·029	·035	·046	·030	·026

* See remark on this line on p. 155.
† Gale *et al.* interferometer measures.
‡ Figures in this row are taken from table 3.
§ Measured in one order only; the other figures in this row are from two orders.

From table 2 are obtained the mean defects:

$$\nu_G - \nu_F = {\cdot}035\,\text{cm.}^{-1} \quad \text{and} \quad \nu_G - \nu_{39} = {\cdot}033\,\text{cm.}^{-1}.$$

Hence it is concluded that there is no important systematic error in our final (ν_F) results, of table 3, unless there is some fundamental error of method.

9. ON THE DIFFERENCE BETWEEN OUR RESULTS AND THOSE OF GALE, MONK AND LEE

From tables 1 to 3 it is clear that there is an appreciable difference between our vacuum wave numbers and those of Gale *et al.* The irregular part of this difference in some cases is mainly due to the difficulty of eliminating completely the interference of neighbouring lines in this closely packed spectrum. But there is evidently a systematic difference of roughly 0·03 cm.$^{-1}$, in the region 15,500 to 16,600 wave numbers, our values being lower. For this region, about 0·015 cm.$^{-1}$ of the difference is accounted for by the different air-dispersion formulae used in the two cases.*

There seems to be no reason to doubt the validity of our results within the limits given. The residual difference of about 0·015 cm.$^{-1}$ (0·006 A.) is difficult to account for.

---

* Gale *et al.* λ_{air} measurements were reduced with the formula of Meggers & Peters (1918), whilst in our case the vacuum wave number of the Cd line was computed using the formula of Sears & Barrell (1939).

10. ON DEDUCING THE STRUCTURES

In this, the minimum amount of attention was paid to what was expected of a line, but in some of the more confused and badly blended lines it was often necessary to have some guidance as to structure besides what was provided by our observations. The structures given may not invariably be a unique interpretation of the observations in such cases, but, coupled with making a tolerably good fit with the ν of Gale *et al.* and a reasonable balance with their intensities and also a fair fit with what might be expected from the stucture and properties of related lines for which the determinations are free from ambiguity, it is believed that the structures are substantially correct in most, if not all, cases. In the relatively few cases where our results show a substantial disagreement with those of Gale *et al.*, it may sometimes be taken as a sign that something is amiss, but this does not always apply, since the poorest fits are mostly concerned with lines that were evidently not properly resolved by Gale *et al.*, and hence probably not accurately measured. Moreover, the intensity distribution in our spectrum and that of Gale *et al.* or of Merton & Barratt (1922) is quite appreciably different. On our plates, as a rule, lines with high rotational quantum number (say $n > 3$) are relatively quite weak compared with the Gale *et al.* value, so that in complex lines involving, for example, one such line and a low rotational line, comparatively large differences in the composite ν may be tolerable.

11. EXAMPLE OF ANALYSIS OF A COMPLEX STRUCTURE

We shall now describe an example of the treatment used in one of these, relatively few, instances where the echelon observations together with the results of Gale *et al.* are insufficient to provide a unique interpretation of complex patterns; so that we are obliged to make use of some empirical and certain theoretical predictions to analyse the patterns. No simple routine can be stated since every case is different, but the following example gives some idea of the kind of treatment.

The example chosen is the following group of lines taken from table 3:

ν_G	intensity	classification	reference letter	approx. $2\nu t$ (order of interference)	
16917·50	(3)	$3s\,^3\Sigma \rightarrow 2p\,^3\Pi,\ 1 \rightarrow 1,\ Q\,1$	X	23273·61	(1)
16916·57	(0)	$3s\,^3\Sigma \rightarrow 2p\,^3\Pi,\ 1 \rightarrow 1,\ Q'\,2$ and ?$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi,\ 2 \rightarrow 2,\ P\,6$	Y	23272·34	
16915·33	(3)	$3s\,^3\Sigma \rightarrow 2p\,^3\Pi,\ 1 \rightarrow 1,\ Q\,3$	Z	23270·63	

The path difference $2t$ corresponding to the position of the intensity maximum is about $1{\cdot}375712_4$ cm., so that the order of interference at the position of the intensity maximum is as given in the right-hand column above. The range of wave numbers in this group is only about 2·2 so that the members appear as a single pattern. Comparing $2\nu t$'s it is clear that overlapping orders will be formed for the lines in X and Y if both are single or if both have similar structures.

Our observations on this group combined measurements on four lines designated $(A+)$, $(?b)$, (B) and (A), where $(A+)$ was evidently the next higher order of (A) and need not be considered further for the present discussion. This leaves:

component	$(?b)$	(B)	(A)
observed intensity	1/3	3	5
fractional part, f_m	$-0{\cdot}223$	$-0{\cdot}438$	$-0{\cdot}681$

(2)

It will be noted that none of these fractional parts corresponds with the fractions of the listed orders of interference (1). Now it is clear from the intensities (3) and (3) assigned respectively to X and Z and from the assignments of intensity (3) and (5) to (B) and (A) respectively that as far as the bulk of the intensity is concerned, (B) and (A) together somehow belong to the lines X and Z. Since the decimal parts of the orders of interference for X and Z are about the same and since Gale *et al.* intensities for these lines are also the same, this suggests that about one-half of (B) and one-half of (A) should be assigned to X and the other halves of each to Z. Assigning weights 3 and 5 to lines (B) and (A) respectively, then

$$-\left\{\frac{3\times 0{\cdot}438+5\times 0{\cdot}681}{8}\right\}=-0{\cdot}59$$

is obtained for the fractional parts of the 'optical centre' of these components, in satisfactory agreement with the decimal parts of the orders of interference for X and Z.

In further support of the above is the fact that X and Z are classified lines, and it is known from results obtained on many rather similar lines ending on $2p\,{}^3\Pi$ that each is likely to appear as a doublet, with components a and b about 0·18 cm.$^{-1}$ apart and intensity ratio $\dfrac{I(b)}{I(a)}=\dfrac{2}{5}$. It therefore appears that X is composed of $\begin{Bmatrix}a_1 \text{ intensity } 5\\ b_1 \text{ intensity } 2\end{Bmatrix}$ and Z is composed of $\begin{Bmatrix}a_2 \text{ intensity } 5\\ b_2 \text{ intensity } 2\end{Bmatrix}$, so that $\dfrac{I(b_1+b_2)}{I(a_1+a_2)}=\dfrac{2}{5}$; this ratio should be $=\dfrac{I(B)}{I(A)}=\dfrac{3}{5}$. This suggests that B still contains some unassigned strength.

Turning now to the line Y, one has the component $(?b)$ and the unassigned intensity of about 1 in (B). It if be supposed that these two constitute the line Y, then $-\left\{\dfrac{1\times 0{\cdot}223+3\times 0{\cdot}438}{4}\right\}=-0{\cdot}38$ is obtained for the fractional part of the 'optical centre', which is in good enough agreement with the decimal part of the corresponding $2\nu t$. Thus to Y is assigned components a_3, b_3, having intensities 1 and 1/3 respectively.

It thus appears that the three observed components $(?b)$, (B) and (A) are the result of the blending of at least six lines of which component (A) has two and (B) has three. The effect on the pattern of the possible $2\rightarrow 2$, $P\,6$ in Y is assumed to be negligible since it would be very weak in our spectrum.

A final check on these assignments is provided by the combinations between the individual components a_1, b_2, a_2, b_2, a_3, b_3 with other lines having final levels of the $2p\,{}^3\Pi$ state. These will be considered in a subsequent paper dealing with the analysis of the lines ending on that state.

In table 3 the observed $\Delta\nu$'s for 16,917·50 and 16,915·33 are both 0·193 cm.$^{-1}$ but it is known from the behaviour of many lines ending on $2p\,{}^3\Pi$ that more probable values are the alternatives also entered in column 3. Our incomplete resolution does not show up these small differences.

12. GENERAL NOTES TO TABLE 3

Columns 1–3. Column 1 gives Gale *et al.* values of ν for the lines followed by their intensity estimates. Column 2 gives the transitions of the classified lines, the notation used being that of Richardson (1934). The abbreviation α is used for the Fulcher system $3p\,{}^3\Pi \rightarrow 2s\,{}^3\Sigma$. In column 3, $\Delta\nu$ gives the separation of the fine-structure components of the lines ending on $2p\,{}^3\Pi$ of column 1 where this is great enough to be measured.

Column 4. In column 4 the notation used indicates the fine-structure components. Lines ending on $2p\,{}^3\Pi$ have two components, but as they usually differ in intensity one may not appear on the photographs.[†] These are always referred to as *a* and *b*, *a* being the higher frequency and also the stronger component; in some cases numerical subscripts are added to these letters. Letters other than *a* and *b* are used to designate structures which have not been allocated as components of a line ending on $2p\,{}^3\Pi$.

Column 5. The numbers in column 5 represent the intensities of the observed fine structures. Where *only one component* is observed in any particular line and there is no reason to suspect others of appreciable strength, the intensity I is always written $I = 1$. If another component is suspected (for example, because our measurements do not agree with those of Gale *et al.*) but has not been observed we write $I = \sim 1$.

[†] On plate 4 the *a* components are always to the right-hand side of the *b* component in the same order of interference. The following examples show the doublet $(a+b)$ near the single-order position, so that the intensity of the images represent approximately the relative intensities of the components: 16132·09 (5), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $2 \rightarrow 2$, $Q\,3$ (the lower part of the image of the *b* component is covered by 16126·19 (10) in this picture): 16603·20 (9), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $Q\,1$: 16710·63 (6), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $P\,3$. (In this instance *a* and *b* are unresolved—$\Delta\nu = 0{\cdot}06$ cm.$^{-1}$—and their intensities are nearly equal. The direct evidence that the line is a doublet is the abnormal width of the pattern.)

Cases in which the doublet is not close to the single-order position are numerous. Owing to the intensity envelope effect the weaker (*b*) component is frequently not seen in the lower (left-hand) of the two orders. Examples are: 16458·03 (5), $3d\,{}^3\Pi_d \rightarrow 2p\,{}^3\Pi$, $3 \rightarrow 3$, $R\,1$, 16654·24 (9), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $0 \rightarrow 0$, $Q\,3$; 16692·00 (5), $3d\,{}^3\Pi_d \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $P\,3$; 16764·12 (4), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $0 \rightarrow 0$, $Q'\,2$; 16854·86 (10), $3d\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $0 \rightarrow 0$, $Q\,1$; 16978·52 (10), $3d\,{}^3\Pi_d \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $R\,1$; 17192·31 (8), $3d\,{}^3\Delta_c \rightarrow 2p\,{}^3\Pi$, $2 \rightarrow 2$, $Q\,3$; 17261·34 (4), $3d\,{}^3\Pi_c \rightarrow 2p\,{}^3\Pi$, $0 \rightarrow 0$, $Q\,3$; 17279·01 (7), $3d\,{}^3\Delta_d \rightarrow 2p\,{}^3\Pi$, $2 \rightarrow 2$, $R\,1$ (here the weaker *b* component also can be seen in the order on the right where the *a* component is missing): 17316·56 (5), $3d\,{}^3\Delta_c \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $Q'\,2$; 17426·25 (10), $3d\,{}^3\Delta_d \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $R\,1$.

The symbol ~ is also used to denote intensities less accurate than usual *when more than one component* is observed.

When two or more components of any single line are observed, the intensity figures are intended to be rough values of *intensity ratio*. The figures are rough, since (except where otherwise indicated) they are derived from ordinary visual estimates, but a correction has always been applied for inaccuracy caused by components occurring at different positions on the echelon intensity envelope.

The intensity-ratio system has been extended to include all components of two or more *different* lines when an asterisk* is placed beside the intensity figure. If this system is applied to two different groups of close lines which occur consecutively in the table, the members of the second group are distinguished from those of the first by having two asterisks **. The intensities of members of each group are relatively correct, but the figures are not consistent in passing from one group to another.

Column 6. Column 6 gives the decimal part of our ν values for the structures, the integral parts being the same as for the corresponding figures of column 1 except where a change is indicated by the insertion of the units figures in column 6. It should be remembered that our ν is usually about 0·03 cm.$^{-1}$ less than that of Gale *et al.*

Column 7. We believe that the figures given as tolerances in column 7 correspond to about twice the probable error (which can seldom be evaluated owing to indeterminate effects caused by blends).

In such a complex spectrum as H_2, using a wide slit, and with only the coarse dispersion of a Hilger E 1 spectrograph available to separate the lines, many cases of overlapping patterns and overlapping orders of different lines necessarily occur (quite apart from the numerous blends known to be present in the lines in Gale *et al.* list in addition to those explicitly mentioned). Knowing the echelon step width it is easy from Gale *et al.* table to predict when such overlapping orders are likely to occur in neighbouring lines. To condense our notes we have not, therefore, always drawn attention to these instances of overlap. However, where they are known to occur or where they might be expected to occur we have allowed a greater tolerance than would have been quoted otherwise.

Ordinarily where there is no confusion the tolerance figures depend on the number of times a line has been measured and on the concordance of the results.

In general, the tolerance figures given are a kind of resultant of the effect of the concordance of the measured values and of the factors referred to in the original notes to table 3 (see below).

Column 8. The numbers in column 8 refer to a series of notes which will be found at the end of table 3.

Our original notes to table 3 are elaborate and are too bulky for publication. They include remarks on (*a*) blending, including predicted wave numbers and relative intensities of the lines of a blend, (*b*) details of our analysis of complex structures, and (*c*) new observational data which could not be tabulated. The residue of these notes referred to in column 8 is mainly concerned with (*c*) and a few of the more outstanding of the other items.

(*a*) Wave numbers of blended lines are predicted by starting with a related line whose wave number is known and whose combination difference with the blended line is also known. The accuracy of this process can vary a great deal. There may be as many as six or more observationally sharp lines all free from blends and all connected with the predicted line by accurately known combinations.

In such a case there should be little difference between the six or more results, and the mean should be of the same kind of accuracy as that of the measurement of any good line in the spectrum.

At the other extreme the only related line available may itself be a blend. In this case one has to calculate the frequency of the constituents one is interested in. This requires a knowledge of the intensities of the two constituents. This can rarely be obtained accurately. Finally, the known combination on which one depends may not be very reliable.

There is obviously scope for much variation in the tolerance of the final result given by these processes, and assessing its value requires delicate judgement.

Predicted wave numbers are always given in square brackets [ ], and unless otherwise stated they are on our scale, i.e. ν_F. It should be remembered that $\nu_G \approx \nu_F + 0{\cdot}03$. Predicted intensities are also given in [ ]. $\frac{I(a)}{I(b)}$ is frequently written $I(a/b)$ for convenience of printing. Similarly, $I\frac{(R\,0)}{(Q\,2)}$ means $\frac{I(R\,0)}{I(Q\,2)}$, etc. If, for example, $I\frac{(R\,0)}{(Q\,2)} = \left[\frac{8K}{3K}\right]$ this means that $I(R\,0) = 8K$ and $I(Q\,2) = 3K$, K referring to the scale of Kapuscinski & Eymers (1929).

(*b*) To cut down additional notes we have frequently used the following scheme to indicate briefly how the various components in a complex pattern (including one or more of Gale *et al.* lines) have been assigned:

As already pointed out, intensities of members of any particular complex pattern are usually asterisked, i.e. they are all on the same scale. In such a group, besides the usual letter for each component of a line (column 4) another letter is quoted in brackets. These bracketed letters indicate the *observed* components, and they may occur more than once amongst the various members of the asterisked group showing how each observed component has been assigned to the different transitions. See, for example, the lines 16,915·33, 16,916·57, 16,917·50 in table 3. In such cases, corresponding bracketed intensities are also given, the unbracketed intensities indicating how the observed intensity has been distributed amongst the various lines. In several such cases there are not enough data to distribute the intensities uniquely from the purely observational material.

TABLE 3

G.M.L. ν cm.$^{-1}$ $I(G)$	transition	$\Delta\nu$ cm.$^{-1}$	components	intensity ratio I	F.R. ν cm.$^{-1}$	tolerance cm.$^{-1}$	notes
17859·74 (10)	α, 3→2, $P3$	—	z	~1	·696	0·01	—
17812·31 (10)	$^3Y \rightarrow 2s\,^3\Sigma$, 1→1, $R1$	—	z	~1	·253	0·01	—
17741·92 (8)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 0→3, $R2$	—	z	1	·909	0·015	—
17676·23 (9)	α, 4→3, $Q1$	—	z	1	·203	0·007	—
17660·45 (1)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 1→5, $R0$	—	z	1	·434	0·04	—
17657·88 (3)	unclassified	—	z	1	·869	0·015	—
17629·46 (1a)	$^1N \rightarrow 2p\,^1\Sigma$, 0→6, $R2$	—	z	~1	·424	0·02	—
17628·90 (4)	$^3Y \rightarrow 2s\,^3\Sigma$, 2→2, $R2$ and α, 4→3, $Q3$	—	y	~1	·80	0·08	—
17616·32 (2)	$3^10 \rightarrow 2p\,^1\Sigma$, 1→6, $P2$ and $3\alpha\,^3\Pi_d \rightarrow 2p\,^3\Pi$, 4→3, $P5$	—	z	~1	·281	0·02	—
17591·22 (4)	?α, 4→3, $Q4$ and	?0·15$_2$	a_1 (b)	3*	·222	0·03	—
	$3d\,^3\Delta_d \rightarrow 2p\,^3\Pi$, 1→1, $R3$		b_1	? ~1*	~ ·07	0·06	
17589·17 (3)	$^3Y \rightarrow 2s\,^3\Sigma$, 1→1, $P2$	—	y (a)	8*	·107	0·02	—
17588·00 (5)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 1→5, $R0$ and		x (c)	??1½ (2*)	8·148	0·05	—
	$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi$, 0→0, $R'1$	0·12$_9$	a_2 (d)	8*	7·962	0·02	
			b_2 (e)	2*	7·833	0·04	
17586·73 (1)	$^1N \rightarrow 2p\,^1\Sigma$, 0→6, $R1$	—	z (c)	??½ (2*)	·694	0·05	—
17572·31 (10)	$3d\,^3\Delta_d \rightarrow 2p\,^3\Pi$, 0→0, $R1$	0·213	a	2·5	·364	0·006	
			b	1·0	·151	0·015	
			p_1	0·7$_5$	·099	0·02	unclassified (1)
17557·05 (5)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 1→5, $R2$	—	z	1	·030	0·015	—
17537·02 (4)	$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi$, 0→0, $Q3$	0·216	a	2·5	7·023	0·03	—
			b	1	6·807	0·03	
17529·00 (5)	$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi$, 1→1, $R2$	—	a	~1	8·992	0·02	
			?b	$<0·2_5$	—	—	not observed
17527·44 (4)	$3d\,^3\Delta_d \rightarrow 2p\,^3\Pi$, 0→0, $Q4$	?0·08	a	?1·5	·413	0·02	—
			b	1			
17508·99 (5)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 0→3, $P4$ and	—	z	??1	·860	0·02	
	$3d\,^3\Delta_d \rightarrow 2p\,^3\Pi$, 1→1, $R'2$						not observed
17497·72 (5)	$^3Y \rightarrow 2s\,^3\Sigma$, 1→1, $P3$	—	z	1	·703	0·006	—
17494·35 (1)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, 1→5, $P2$	—	y	1	·335	0·02	—
17467·13 (6)	$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi$, 0→0, $Q'2$	0·160	a	2·5	·220	0·007	—
(0·19 M.B.)			b	1	·060	0·02	
17451·59 (10)	$3d\,^3\Delta_d \rightarrow 2p\,^3\Pi$, 0→0, $Q2$	0·149	a	P. 86 P. 40 2·5–3 4–5	·593	+0·02 −0·04	sharp (P. = plate)
			b	1 1	·444	0·006	?blend (2)
17441·32 (8)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 0→0, $R2$ and	0·169	a	2·5	·347	0·006	—
	$3d\,^1\Pi_c \rightarrow 2p\,^1\Sigma$, 2→7, $P2$		$b+q$	[0·6+0·4]1	·178	0·01	(3)
17433·50 (1)	$3d\,^3\Delta_c \rightarrow 2p\,^3\Pi$, 2→2, $R'3$	?0·2$_1$	??a_1 (?y)	?1*	·590	0·05	
			?b_1 (x)	? ~1/3 (1)*	·378	0·02	(4)

17431·57	(6)	$3d\,^1\Sigma\to 2p\,^1\Sigma$, 2→7, $R\,4$ and $3d\,^3\Delta_c\to 2p\,^3\Pi$, 1→1, $R'\,1$	—	w a_2	(a)	7*	·559	0·006	(a) sharp
				b_2		—	—	—	b_2 not observed
				z	?(p)	½*	·727	0·04	unclassified
17426·25	(10)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 1→1, $R\,1$	0·190	a		2·5	·276	0·004	—
				b		1	·086	0·006	
17421·59	(2)	$3d\,^1\Sigma\to 2p\,^1\Sigma$, 2→7, $R\,2$ and $^1N\to 2p\,^1\Sigma$, 0→6, $P\,4$	—	z		~1	·566	0·015	—
17411·21	(5)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 1→1, $Q\,6$	~?0·08	?a ?b		?1 1	·183	0·02	unresolved
17404·90	(2)	$3d\,^1\Pi_d\to 2p\,^1\Sigma$, 2→7, $Q\,2$	—	z		1	·872	0·015	—
17369·32	(5)	$3d\,^3\Pi_c\to 2p\,^3\Pi$, 0→0, $R'\,1$	—	a		~1	·307	0·006	—
				b		≤ 1/5	—	—	not observed
17364·30	(5)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 1→1, $Q\,4$	0·134	a		1	·350	0·02	—
				b		?1	·216	0·02	
17357·64	(6)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 1→1, $Q\,3$	0·200	a		2·5	·693	0·02	—
				b		1	·493	0·02	
				??x		0·25	·828 or ·101	0·04	
17355·14	(6)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 2→2, $R\,2$ and $^1Q\to 2p\,^1\Sigma$, 0→3, $P\,4$	~0·13	a		4	~ ·17	0·05	—
				b		1	~ ·04	0·04	
				??q		≤0·5	~4·9	0·1	
17348·12	(3)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 2→2, $R'\,2$	?0·138	a		2·5	·144	0·02	—
				?b		?1	·006	0·03	
17342·83	(1)	$3\,^1K\to 2p\,^1\Sigma$, 1→5, $P\,2$	—	?z		1·	·810	0·03	—
17316·56	(5)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 1→1, $Q'\,2$	0·161	a		2 5	·579	0·008	—
				b		1	·418	0·015	
17312·48	(3)	α, 5→4, $Q\,1$	—	?z		—	·423	0·07	—
17311·08	(9)	$^3Y\to 2s\,^3\Sigma$, 1→1, $P\,5$ and $3d\,^3\Delta_d\to 2p\,^3\Pi$, 1→1, $Q\,2$	?0·138	a		~6	·078	0·02	—
				?b		?1	0·940	+0·01	
								−0·05	ν probably too high
17299·30	(4)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 0→0, $P\,4$	0·08	a b		1 1·5	·267	0·03	unresolved
17296·19	(2h)	$3\,^1K\to 2p\,^1\Sigma$, 2→7, $R\,2$	—	x'		1*	·269	0·03	—
				y		3*	·068	0·03	—
17294·13	(1)	α, 5→4, $Q\,2$	—	x		3*	·088	0·05	—
17287·27	(2)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 0→0, $P'\,3$ and $3d\,^3\Delta_d\to 2p\,^3\Pi$, 3→3, $R'\,4$	—	z		—	·252	0·03	— —
17280·68	(3)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 2→2, $R'\,1$	?0·096	a		6	·661	0·02	—
				?b		?1	·565	0·05	
17279·01	(7)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 2→2, $R\,1$	0·185	a		2·5	9·015	0·007	—
				b		1	8·830	0·004	
17271·56	(4)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 0→0, $P\,3$	0·144	a		2·5	·552	0·01	—
				b		1	·408	0·02	
17261·34	(0)	$3d\,^3\Pi_c\to 2p\,^3\Pi$, 0→0, $Q\,3$	0·151	a		2·5	·310	0·015	(5)
				b		1	·159	0·015	
17252·65	(2)	?$3d\,^3\Pi_c\to 2p\,^3\Pi$, 0→0, $Q\,5$	?0·15	a		2	·639	0·03	
				?b		? ~1	·49	0·05	(6)

Table 3 (*continued*)

G.M.L. ν cm.$^{-1}$ $I(G)$	transition	$\Delta\nu$ cm.$^{-1}$	components	intensity ratio I	F.R. ν cm.$^{-1}$	tolerance cm.$^{-1}$	notes
17248·55 (1)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 0→0, $Q'2$	—	a	2·5	·522	0·02	—
			b	[1]	[·34]	—	not observed
17218·50 (8)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 1→1, $R2$ and	0·118	a	~2	·494	0·015	(7)
	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 2→2, $Q'4$		b	1	·376	0·03	
17212·01 (1)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 0→0, $Q1$ and	?0·185	?a	2	·012	0·04	—
	?β, 4→6, $Q1$		?b	?1	1·827	0·04	
17199·28 (10)	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 0→0, $R3$ and	0·160	a_1	2·5	·388	0·005	—
			b_1 a_2 (a_2b_1)	[1] [2·5] 3·5	·228	0·008	sharp unresolved
	0→0, $R1$	0·216	b_2	1	·012	0·005	
17192·31 (6)	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 2→2, $Q3$ and	0·150	a	2·5	·280	0·006	—
	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 0→0, $P7$		b	1	·130	0·01	$P7$ not observed
17186·00 (2)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 3→3, $R'2$	??0·136	??a	2·5	·018	0·08	—
			??b	??1	5·882	0·08	
17184·46 (3)	$3d\,\Delta_d \to 2p\,^3\Pi$, 2→2, $Q'3$	?0·108	a	?3	·428	0·05	—
			b	1	·320	0·05	
17171·27 (4)	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 2→2, $Q'2$	0·175	a	3	·290	0·03	—
			b	~1	·115	0·01	
			?x	0·3	~ ·51	0·04	?x unclassified
17169·22 (10)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 2→2, $Q2$	0·145	a	5	·232	0·01	—
			b	~1	·087	0·03	
17144·93 (4)	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 1→1, $P'3$		p	1–2	·997	0·02	unclassified
		0·103	a	4	·853	0·02	
			?b	1	·75	0·05	(8)
17139·77 (5)	$3s\,^3\Sigma \to 2p\,^3\Pi$, 1→1, $R'3$ and	0·172	a (x)	2·5*	·771	0·01	—
	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 1→1, $P3$		b (y)	?1 (3)*	·599	0·03	
17138·90 (4h)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 1→1, $R'1$	—	a (y)	?2 (3)*	·872	0·02	—
			b	—	—	—	not observed
17130·78 (3a) (M.B. 0·36)	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 1→1, $P4$ and		a_1	[1]	[~ ·78]	—	—
			b_1	[~1]	[~ ·66]	—	
	?$3d\,^3\Delta_c \to 2p\,^3\Pi$, 3→3, $R'1$	0·176	a	5	·276	0·02	
			b	~1	·10	0·06	
17129·89 (8)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 3→3, $R1$	0·182	a	2·5	·917	0·008	—
			b	1	·735	0·005	
17125·07 (2)	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 0→0, $Q'1$	?? ~0·20	a	5	·035	0·05	(9)
			??b	?1	—	—	
17091·29 (8)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 0→0, $P2$	0·172	a	4	·277	0·01	—
			b	1	·105	0·01	
17086·67 (2)	$3s\,^3\Sigma \to 2p\,^3\Pi$, 1→1, $R2$	? ~$0{\cdot}2_{25}$	?a	?3	·70	0·06	—
			b	1	·475	0·02	
17084·56 (2)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 1→1, $P5$	? ~0·08	?a	?1	·56	0·06	—
			b	1	·48	0·03	

17068·60 (1)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 0→0, $P'3$	? ~0·10	a		?1	·557	0·04	—
			b		1			
17060·68 (5)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 4→4, $R3$	~0·13	a	(a)	1–2	·668	0·03	broad and hazy
			b		1			
17047·15 (4)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 1→1, $Q3$	0·170	a		2·5	·189	0·005	—
			b		1	·019	0·01	
[17033·66]	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 3→3, $Q4$	0·167	a	(y)	? ~½* (1½)	·783	0·05	—
			b	(x)	½*	·616	0·04	
17033·21 (4)	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 3→3, $Q3$	0·160	a	(z)	2½*	·217	0·02	
			b	(y)	?1 (1½)*	·057	0·04	(10)
17026·36 (2)	?$3d\,^3\Delta_c \to 2p\,^3\Pi$, 3→3, $Q'2$	0·179	a_1	(a)	2½**	·368	0·02	—
			b_1	(b)	?1**	·189	0·06	ν probably low (11)
17025·41 (6)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 3→3, $Q2$	?0·16$_6$	a_2	(b) (q)	2 or ?5**	·40$_5$	0·05	—
			b_2	(p)	1**	·23$_9$	0·08	(11)
17023·67 (1)	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 0→0, ?$P4$ and	0·09	a	(?o)	?1**	·63	0·02	—
			b_3	(?m)	?1**	·54	0·04	(11)
		[~0·16]	a_4		[2·5]	[~ ·86]	—	
	1→1 do. $Q'2$		b_4		[1]	[~ ·70]	—	
17006·43 (10)	unclassified		x	(A)	3	·564	0·005	—
	unclassified		?y	(B)	3 or 4	·350	0·015	sharp unresolved
	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 2→2, $P3$	0·166	a_1		? ~2½		0·02	
			b_1	(p)	1	·184	0·02	
17004·40 (3a)	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 0→0, $P'2$ and	—	z	(q)	—	·583	0·02	
	$3s\,^3\Sigma \to 2p\,^3\Pi$, 2→2, $R2$		??x		—	·20	0·06	(12)
16996·32 (1a)	$^1M \to 2p\,^1\Sigma$, 0→5, $R1$	—	z		1	·227	0·01	—
			?x		~1	~ ·40	0·1	
16990·71 (6)*	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 1→1, $R3$		p	(a)	~7*	·955	0·02	unclassified
(90·91 M.B.)		? ~0·21	a_1	(b)	~3*	·711	0·05	broad (13)
			b_1	(?x)	1*	? ~ ·50	0·05	not measured
16988·72 (8)	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 1→1, $R'2$ and	~0·22$_9$	a_1	(y)	2*	·864	0·03	—
			b_1	(A)	? ~1 (9)*	·635	0·04	(A) fairly sharp
	$3d\,^3\Pi_c \to 2p\,^3\Pi$, 1→1, $Q1$	0·220	a_2		? ~8	·635	0·02	unresolved (14)
			b_2	(B)	4*	·415	0·02	?rather broad
16980·89 (1)	$3s\,^3\Sigma \to 2p\,^3\Pi$, 0→0, $Q3$ and	—	?x		~¾	1·01	0·05	—
			z		1	·899	0·02	
	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 4→4, $R1$		?y		~¾	·71	0·05	
16978·52 (10)	$3d\,^3\Pi_d \to 2p\,^3\Pi$, 1→1, $R1$	0·205	a		2·5	·548	0·005	—
			b		1	·343	0·006	
16976·03 (5)	unclassified	—	—	(p)	½	·105	0·02	—
	$3d\,^3\Delta_c \to 2p\,^3\Pi$, 2→2, $P4$,	prob. broad	a	(x)	1	5·932	0·02	
	[$^1M \to 2p\,^1\Sigma$, 0→5, $Q5$] and		$b+r$			ν from		
	?0→0, $3s\,^3\Sigma \to 2p\,^3\Pi$, $Q5$					P. 40 5·942		P.=plate.
16962·63 (1)	α, 6→5, $Q1$	—	z		1	·606	0·005	—
16940·99 (1)	$3p\,^3\Sigma \to 2s\,^3\Sigma$, 6→2, $R1$	—	z		?1	·936	0·015	—
			??x		??¼	1·00	0·05	
16934·41 (1)	$3d\,^3\Delta_d \to 2p\,^3\Pi$, 2→2, $P5$	0·14	?a	(y)	?2 (3)*	·463	0·04	—
			?b	(x)	2*	·323	0·04	

Table 3 (*continued*)

G.M.L. ν cm.$^{-1}$	$I(G)$	transition	$\Delta\nu$ cm.$^{-1}$	components		intensity ratio I	F.R. ν cm.$^{-1}$	tolerance cm.$^{-1}$	notes
16930·14	(0)	$?3s\,^3\Sigma\to 2p\,^3\Pi$, 0→0, $P'1$	—	r	(y)	?1 (3)*	~ ·102	0·05	—
16917·50	(3)	$3s\,^3\Sigma\to 2p\,^3\Pi$, 1→1, $Q1$	prob. ~0·20 obs. 0·193	a_1	(A)	2½ (5)**	·539	0·02	—
				b_1	(B)	?1 (3)**	·346	0·02	(15)
16916·57	(0)	$3s\,^3\Sigma\to 2p\,^3\Pi$, 1→1, $Q'2$ and $?3d\,^3\Delta_c\to 2p\,^3\Pi$, 2→2, $P6$	? ~$0{\cdot}16_2$	$?a$	(B)	? ~1 (3)**	·628	0·04	—
				b		~1/3**	·466	0·02	(15)
16915·33	(3)	$3s\,^3\Sigma\to 2p\,^3\Pi$, 1→1, $Q3$	prob. ~0·18 obs. 0·193	a_2	(A)	2½ (5)**	·358	0·02	—
				b_2	(B)	?1 (3)**	·165	0·02	(15)
16899·76	(0)	$?3d\,^3\Pi_d\to 2p\,^3\Pi$, 1→1, $Q'1$ and Δ_d, 2→2, $P'6$	—	a_1	(b)	(1)*	$\cdot 53_1$	0·04	—
				b_1		—	—	—	not observed
16888·48	(5)	$3d\,^3\Pi_d\to 2p\,^3\Pi$, 0→0, $P3$	0·177	a_2	(B)	(6)*	·432	0·03	—
				b_2	(a)	(2½)*	·255	0·04	
16897·19	(5)	$3d\,^3\Pi_c\to 2p\,^3\Pi$, 2→2, $R2$	—	a_3	(A)	(4)*	·226	0·02	—
				b_3		—	—	—	not observed
16892·70	(5)	unclassified	0·194	—	(x)	2·5	·732	0·005	—
				—	(y)	1	·538	0·01	
16885·83	(5)	unclassified	—	p	(x)	? ~3 (4)**	·800	0·01	—
16884·94	(4)	$3d\,^3\Sigma\to 2p\,^3\Pi$, 0→0, $R'1$ and	~$0{\cdot}14_6$	a	(y)	? ~4½ (5)**	·896	0·01	—
				$?b$		1**	·75	0·04	
		$3d\,^3\Delta_c\to 2p\,^3\Pi$, 4→4, $Q3$		a	(x)	? ~1 (4)**	5·075	0·05	—
				b	(y)	? ~½(5)**	4·896	0·05	
				$?w$		1/6**	4·549 or 5·277 or 6·003	0·05	unclassified
16873·46	(9)	$3d\,^3\Pi_c\to 2p\,^3\Pi$, 1→1, $P2$	0·182	a		5	·453	0·015	—
				b		1	·271	0·01	
16870·70	(2)	$3d\,^3\Delta_d\to 2p\,^3\Pi$, 3→3, $P3$ and $3s\,^3\Sigma\to 2p\,^3\Pi$, 0→0, $P2$ and $?3d\,^3\Delta_c\to 2p\,^3\Pi$, 3→3, $P'3$	~$0{\cdot}15_1$	a		~2	·742	0·015	—
				b a_1	(b')	1	·591	0·02	
				b_1		—	—	—	b_1 not observed
16854·86	(10)	$3d\,^3\Sigma\to 2p\,^3\Pi$, 0→0, $Q1$	0·194	a		2·5	·870	0·004	—
				b		1	·676	0·005	
16852·22	(3)	$3d\,^3\Pi_c\to 2p\,^3\Pi$, 1→1, $P'3$	0·166	a		3 (or 4)	·174	0·005	—
				$??b$		1	·008	0·03	
16841·63	(4)	$3s\,^3\Sigma\to 2p\,^3\Pi$, 2→2, $Q1$, $Q'2$, $Q3$ and $Q'4$	—	$?p$		~1	·793	0·03	
				y		9	·608	0·005	(16)
				x		2 or 3	·261	0·015	
16834·27	(10)	$3d\,^3\Sigma\to 2p\,^3\Pi$, 0→0, $R2$	0·114	a	(l)	8	·273	0·015	
				b	(m)	2	·159	0·03	(17)
				—	(n)	1	·052	0·05	unclassified
16827·14	(2)	$3d\,^3\Delta_c\to 2p\,^3\Pi$, 3→3, $P4$	prob. broad	a_1	(x)	(4)*	·07	0·05	(18)

16825·98 (1)	$3d^3\Delta_d\rightarrow 2p^3\Pi$, $3\rightarrow 3$, $P'4$ and	?0·11	?a_2	(x)	?1 (4)*	6·027	0·03	(18)
	?$3d^1\Pi_d\rightarrow 2p^1\Sigma$, $0\rightarrow 4$, $P3$		??b_2		(<1)*	5·917	0·05	
16824·76 (3)	unclassified	—	(y)		(4)*	·757	0·03	(18)
16822·00 (2)	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $R'1$	? ~0·17	?a_3	(r)	(2½)*	1·95	0·08	(18)
			?b_3	(q)	(?1)*	1·78	0·08	
16820·29 (2)	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $1\rightarrow 1$, $P4$	?broad ~0·12	a_4 b_4	(p)	(2)*	~ ·24	0·06	(18)
16809·70 (6)	α, $0\rightarrow 0$, $R3$	—	z		1	·662	0·005	—
16807·74 (1)	$3s^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $P'3$	?? ~0·17	a		~3	·694	0·01	—
			??b		??1	·52	0·04	
16802·37 (10)	$3s^3\Sigma\rightarrow 2p^3\Pi$, $1\rightarrow 1$, $P2$	0·15	a		4–6	·361	0·015	(19)
			b		1	·211	0·02	
		?broad	x		1/5	·004 or ·731	0·02	unclassified
16774·97 (4)	$3d^1\Pi_c\rightarrow 2p^1\Sigma$, $0\rightarrow 4$, $P2$ and	—	p	(a)	? ~1 (4)*	5·097	0·05	(20)
	$3d^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $R'3$ and	~0·12	a_1	(b)	~(4)*	4·870	0·02	
	$2\rightarrow 0$, $3^1K\rightarrow 2p^1\Pi$, $Q2$		b_1	(b')	~(1)*	4·75	0·05	
16774·42 (6)	α, $0\rightarrow 0$, $R2$	—	q	(a)	? ~3(4)*	·371	0·03	—
16764·12 (4)	$3d^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $Q'2$	0·189	a		2·5	·148	0·008	(21)
			b		1	3·959	0·005	
16753·44 (4)	$3d^3\Pi_d\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $R3$	0·154	a		3	·450	0·005	(22)
			b + ?x		2	·296	0·015	
16734·20 (4)	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $Q3$ and	?$0·17_0$	?a_1		~1*	$·32_5$	0·05	not measured
			b_1	(r)	1*	$·15_5$	0·04	
	$3d^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $P2$	? ~$0·14_0$	a_2 ?b_2		~½*	$·01_5$	0·05	
16730·55 (10)	$3d^3\Pi_d\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $R1$ and	$0·18_9$	a		~2½*	·709	0·01	—
	α, $0\rightarrow 0$, $R1$		b x	(s)	?1 ?8 (9)*	·520	0·04 0·01	sharp unresolved
			p		1*	·38	0·04	
16710·63 (6)	$3d^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $R4$	0·06	a		~1	·599	0·006	unresolved
			b		1			
16692·00 (5)	$3d^3\Pi_d\rightarrow 2p^3\Pi$, $1\rightarrow 1$, $P3$ and	0·146	a		~2·5	·032	0·01	—
	$3d^1\Sigma\rightarrow 2p^1\Sigma$, $2\rightarrow 0$, $R3$		b		1	1·886	0·015	
16688·40 (4)	$3s^3\Sigma\rightarrow 2p^3\Pi$, $1\rightarrow 1$, $P4$ and	—	—	(p)	?3 (4)*	·373	0·02	—
	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $4\rightarrow 4$, $P4$		—	(y)	1*	7·983 or 8·710	0·01	
16684·88 (1)	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $Q1$	? ~0·17	?a	(x)	3*	·912	0·01	—
			?b	(p)	? ~1 (4)*	·738	0·04	
16678·57 (10)	α, $0\rightarrow 0$, $R0$ and	—	y		~1	·536	0·008	—
	$3d^3\Pi_c\rightarrow 2p^3\Pi$, $3\rightarrow 3$, $R4$		x		~1/20	·59	0·04	
16654·24 (9)	$3d^3\Sigma\rightarrow 2p^3\Pi$, $0\rightarrow 0$, $Q3$	0·173	a		2·5	·253	0·006	—
			b		1	·080	0·006	
16648·40 (2)	$3d^3\Pi_d\rightarrow 2p^3\Pi$, $2\rightarrow 2$, $Q'1$	—	a		1	·411	0·02	—
			b		$\ngtr$ 1/4	—	—	not observed
16642·14 (2)	$3p^3\Sigma\rightarrow 2s^3\Sigma$, $6\rightarrow 2$, $P3$	—	z		1	·099	0·015	—
16630·47 (3)	$3d^3\Sigma\rightarrow 2p^3\Pi$, $1\rightarrow 1$, $R'1$	—	a		1	·468	0·015	—
			b		$\ngtr$ 1/5	—	—	not observed

Table 3 (*continued*)

G.M.L. ν cm.$^{-1}$ $I(G)$	transition	$\Delta\nu$ cm.$^{-1}$	components	intensity ratio I	F.R. ν cm.$^{-1}$	tolerance cm.$^{-1}$	notes
16627·29 (1)	α, 7→6, $Q\,1$	—	z	1	·244	0·02	—
16611·43 (10)	$3p\,^3\Pi_c \rightarrow 2s\,^3\Sigma$, 0→0, $Q\,1$	—	z	1	·399	0·003	—
16603·20 (9)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $Q\,1$ and	0·196	a	2·5	·247	0·006	—
	?$3s\,^3\Sigma \rightarrow 2p\,^3\Pi$, 0→0, $P\,6$		b	1	·051	0·008	
			x	$\frac{1}{2}$	·374	0·015	unclassified
16596·37 (10)	α, 0→0, $Q\,2$	—	z	1	·347	0·004	(23)
16584·72 (10)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $R\,2$	0·118	a	5	·729	0·01	—
			b	1	·611	0·02	
16579·03 (2)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 3→3, $R\,2$ and	? ~0·18	?a_1	? ~1$\frac{1}{2}$	8·97$_7$	0·04	—
	$3s\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $P\,6$		b_1 + ?x (b)	?1/4 + ?3/4 (1)	8·797	0·025	
16575·12 (5)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 2→2, $P\,2$	~0·16$_7$	a	1	·137	0·02	—
			b	$\frac{1}{4}$	4·97	0·05	
16573·94 (10)	α, 0→0, $Q\,3$	—	z	1	·913	0·008	—
16548·90 (3)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 2→2, $P'\,3$ and	? ~0·14	a (a)	~1	·899	0·02	unresolved
	?0→4, $3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, $R\,3$		b	~1			
16544·31 (2)	α, 0→0, $Q\,4$	—	z	1	·278	0·03	—
16538·98 (2)	$3d\,^3\Pi_d \rightarrow 2p\,^3\Pi$, 2→2, $P'\,2$	~0·11	a	~3	·99	0·04	—
			b	1	·88	0·06	
16530·23 (5)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $R'\,3$	~0·12	a	3 or 4	·234	0·03	fuzzy
			?b (p)	~1	·11	0·04	
16517·89 (6)			— (a_2)	~3	8·158	0·03	unclassified
	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 2→2, $P\,4$	0·118	a (b_2)	~2·5	8·024	0·02	(24)
			b (a_1)	~3	7·906	0·02	
			— (b_1)	1	7·794	0·02	unclassified
16515·45 (5)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $Q'\,2$	0·179	a	2·5	·488	0·006	—
			b	1	·309	0·008	
16511·46 (1)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, 3→3, $R'\,1$	~0·15	a	3	·470	0·03	—
			?b	?1	·32	0·05	
16507·71 (2)	α, 0→0, $Q\,5$	—	z	?1	·69	0·05	—
16504·37 (0)	$3\,^10 \rightarrow 2p\,^1\Sigma$, 1→7, $P\,2$	? broad	z	1	·320	0·06	—
16488·15 (8)		—	— (?a_3)	$\frac{1}{2}$–1	·229	0·02	unclassified
			— (a_2)	3	·158	0·01	unclassified
	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $P\,2$	0·10	a_1	1	·056	0·03	(25)
			b_1	~1/3	7·95	0·04	
16489·97 (1)	unclassified	—	x	~1	·945	0·02	—
16479·06 (7)	α, 0→0, $P\,2$	—	z	1	·052	0·01	—
16476·10 (7)	α, 1→1, $R\,2$ and		q (p)	?8 (9)	·057	0·02	—
	$3d\,^3\Pi_d \rightarrow 2p\,^3\Pi$, 3→3, $R\,3$	0·169	a (x)	(2·5)	·226	0·03	
			b (p)	?1 (9)	·057	0·05	
16469·93 (8)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 1→1, $R\,4$	0·10$_0$	a	4	·997	0·02	(26)
			b	5	·897	0·02	
			?x	?$\frac{1}{2}$	·84	0·04	unclassified

16465·55	(2)	$3d\,^3\Pi_d \rightarrow 2p\,^3\Pi$, $3\rightarrow 3$, $R'2$	? $\sim 0{\cdot}17_4$	?a		?1	·654	0·04	—
				?b		1	·480	0·02	
16458·03	(5)	$3d\,^3\Pi_d \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $P3$	0·154	a		2·5	·040	0·004	—
				b		1	7·886	0·006	
?16458·03	(5)	—	—	?x		—	7·702	0·04	(27)
16445·80	(3)	$3d\,^3\Pi_d \rightarrow 2p\,^3\Pi$, $3\rightarrow 3$, $R1$	$\sim 0{\cdot}12_4$	a		~2·5	·823	0·05	—
				b		1	·699	0·04	
16440·72	(10)	α, $1\rightarrow 1$, $R1$ and	—	y		~1	·693	0·003	sharp symmetrical
		$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, $1\rightarrow 6$, $R0$		x		1/30	·385	0·03	
16430·81	(0)	unclassified	—				·764	0·04	
16427·90	(00)	unclassified	—	p		—	? ·856	0·06	—
16425·72	(0)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, $1\rightarrow 6$, $R1$	—				·676	0·04	
16413·37	(7)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $1\rightarrow 1$, $Q3$ and	$0{\cdot}18_2$	a		2·5	·425	0·025	—
		??$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, $1\rightarrow 6$, $R2$		b		1	·243	0·02	
				x		½	·359	0·03	unclassified
				y	(l)	(~1)	2·99	0·04	also 16140·14
16410·14	(0)	$3d\,^1\Sigma \rightarrow 2p\,^1\Sigma$, $1\rightarrow 6$, $R3$	—	z	(l)	(~1)	·086	0·02	also 16413·37
16405·53	(0a)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $1\rightarrow 1$, $R'5$	—	?p		—	·469	0·04	—
16399·79	(10)	α, $0\rightarrow 0$, $P3$	—	z		1	·777	0·005	—
16393·70	(9)	α, $1\rightarrow 1$, $R0$	—	z		~1	·642	0·005	—
16388·32	(1)	$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, $3\rightarrow 3$, $Q1$	~0·21	a		2·5	·366	0·02	—
				?b		~1	·155	0·03	
16344·13	(1)	$^1N \rightarrow 2p\,^1\Sigma$, $0\rightarrow 7$, $P2$	—	—	(z, z_1)	1	·109	0·03	—
16332·11	(1)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $R'1$	—	a_1		—	·141	0·04	—
16330·59	(10)	α, $1\rightarrow 1$, $Q1$		z		~1	·576	0·006	(28)
			?$0{\cdot}16_3$	a	(y)	~1/15	1·087	0·03	unclassified
				b	(x)	~1/40	·924	0·03	
16316·05	(6)	α, $1\rightarrow 1$, $Q2$	—	z		~1	·019	0·006	—
16314·99	(4)	α, $0\rightarrow 0$, $P4$	—	z		~1	·941	0·006	—
16297·25	(5)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $Q1$	$0{\cdot}15_7$	a	(r)	?3*	·294	0·03	—
				b	(q)	?1(3)*	·137	0·05	
16295·02	(3)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $R2$ and	~0·16	?a_1	(q)	?2 (3)*	5·03	0·05	—
		do. $0\rightarrow 0$, $Q'6$		b_1	(p)	½–1*	4·87	0·05	
16294·37	(10)	α, $1\rightarrow 1$, $Q3$	—	z		1	·345	0·02	—
16265·75	(3)	α, $1\rightarrow 1$, $Q4$	—	z		1	·739	0·02	—
16251·79	(4)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, $4\rightarrow 1$, $P3$ and	—	??p		1	·888	0·04	—
		$3d\,^3\Pi_c \rightarrow 2p\,^3\Pi$, $3\rightarrow 3$, $P'3$	$0{\cdot}14_1$	a_1		4	·820	0·02	
				?b_1		1	·679	0·05	
16240·81	(5)	α, $2\rightarrow 2$, $R3$	—	z		~1	·782	0·006	—
16225·06	(8)	α, $0\rightarrow 0$, $P5$	—	z		~1	·035	0·015	—
16222·72	(2)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $Q'2$	0·168	a		2·5	·743	0·025	—
				b		1	·575	0·015	
16208·92	(4)	α, $2\rightarrow 2$, $R2$	—	z		1	·911	0·006	—
16203·93	(7)	α, $1\rightarrow 1$, $P2$	—	z		1	·906	0·006	—
16192·25	(6)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $2\rightarrow 2$, $R4$	$\sim 0{\cdot}12_1$	a_1		? ~2	·312	0·03	—
				b_1		2	·191	0·01	
				c		¼	·42	0·04	unclassified

Table 3 (*continued*)

G.M.L. ν cm.$^{-1}$ $I(G)$	transition	$\Delta\nu$ cm.$^{-1}$	components	intensity ratio I	F.R. ν cm.$^{-1}$	tolerance cm.$^{-1}$	notes
16187·50 (1)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 2\rightarrow2,\ P\,2$	?$0{\cdot}17_3$	a_2 (?y)	$\sim\frac{1}{2}$*	·556	0·03	—
			b_2 (a)	? $\sim\frac{1}{3}$ (3)*	·383	0·04	
16186·63 (3)	$3d\,^3\Pi_a\rightarrow2p\,^3\Pi,\ 3\rightarrow3,\ P\,3$	$0{\cdot}13_2$	a_3 (a)	? ~3 (3)*	·656	0·03	—
			b_3 (b)	~1 (3)*	·524	0·03	
16168·95 (10)	$\alpha,\ 2\rightarrow2,\ R\,1$	—	z	1	·915	0·004	—
16132·09 (5)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 2\rightarrow2,\ Q\,3$	0·166	a	2·5	·134	0·01	—
			b	1	1·968	0·015	
16126·19 (10)	$\alpha,\ 1\rightarrow1,\ P\,3$	—	z	1	·149	0·005	—
16121·52 (8)	$\alpha,\ 2\rightarrow2,\ R\,0$	—	z	1	·492	0·004	—
16060·31 (10)	$\alpha,\ 2\rightarrow2,\ Q\,1$	—	z	1	·293	0·005	—
16046·28 (7)	$\alpha,\ 2\rightarrow2,\ Q\,2$	—	z	1	·252	0·006	
16039·18 (5)	$\alpha,\ 1\rightarrow1,\ P\,4$	—	z	?1	·183	0·01	(29)
16032·09 (1a)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 2\rightarrow2,\ Q'\,4$	—	{?a {?b (x)	~1*	·134	0·03	(30)
16031·77 (2a)	$\alpha,\ 0\rightarrow0,\ P\,7$	—	z	3*	·686	0·008	(31)
16025·36 (10)	$\alpha,\ 2\rightarrow2,\ Q\,3$	—	z	1	·336	0·005	—
15997·73 (2)	$\alpha,\ 2\rightarrow2,\ Q\,4$	—	?z	1	·746	0·02	—
15963·63 (2)	$\alpha,\ 2\rightarrow2,\ Q\,5$	—	?z	1	·651	0·02	—
15949·72 (3)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 3\rightarrow3,\ R\,2$	$\sim0{\cdot}13_5$	{a	3	·767	0·03	—
			{b	2	·632	0·03	
15943·20 (3)	$\alpha,\ 3\rightarrow3,\ R\,2$	—	z	1	·179	0·008	—
15941·22 (5)	$\alpha,\ 2\rightarrow2,\ P\,2$	—	z	1	·216	0·006	?asymmetric
15933·16 (0a)	$^1X\rightarrow2p\,^1\Sigma,\ 3\rightarrow1,\ R\,2$	—	?z	?1	·161	0·06	—
15931·94 (2)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 3\rightarrow3,\ Q\,1$	0·166	a	2·5	2·015	0·008	—
			b	1	·849	0·02	
15927·80 (1)	$^1X\rightarrow2p\,^1\Sigma,\ 3\rightarrow1,\ R\,0$ and	—	z	?1	·795	0·05	—
	$3p\,^1\Sigma\rightarrow2s\,^1\Sigma,\ 1\rightarrow1,\ Q\,1$						
15926·52 (3h)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 2\rightarrow2,\ Q\,5$	—	?p	?1	·479	0·05	—
15923·44 (0) M.B.	$^3H\rightarrow2p\,^3\Pi,\ 0\rightarrow3,\ R\,3$	—	?q	?1	·289	0·05	—
15905·53 (10)	$\alpha,\ 3\rightarrow3,\ R\,1$	—	z	1	·508	0·004	—
15872·20 (1)	?$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 3\rightarrow3,\ Q'\,2$ and	—	?x	?1	·101	0·02	—
	?$3d\,^3\Pi_c\rightarrow2p\,^3\Pi,\ 4\rightarrow4,\ P\,4$ and						
15872·08 (0) M.B.	?$3p\,^1\Pi\rightarrow2s\,^1\Sigma,\ 2\rightarrow1,\ R\,3$	—	—	—	—	—	—
15870·10 (10)	$\alpha,\ 2\rightarrow2,\ P\,3$	—	z	1	·063	0·006	—
15859·86 (5)	$\alpha,\ 3\rightarrow3,\ R\,0$	—	z	1	·849	0·006	—
15827·48 (1)	$3d\,^3\Sigma\rightarrow2p\,^3\Pi,\ 3\rightarrow3,\ P\,2$ (or $P\,2ab$)	—	a	?1	·495	0·03	b not observed
15817·47 (3)	$^1X\rightarrow2p\,^1\Sigma,\ 4\rightarrow1,\ P\,2$ and	—	z	?1	·402	0·01	—
	$4d\,^1\Sigma\rightarrow2p\,^1\Pi,\ 0\rightarrow1,\ R\,3$ and						
	$3p\,^1\Pi\rightarrow2s\,^1\Sigma,\ 1\rightarrow0,\ Q\,5$						

15802·72 (0a)	unclassified	—	y	3*	·728	0·02	
			?x	1*	2·615 or 1·882	0·04	—
15800·76 (10)	α, 3→3, $Q1$ and	—	z	~40*	·743	0·008	—
	?$3p\,^1\Pi \rightarrow 2s\,^1\Sigma$, 1→0, $P2$		?x	1*	1·155	0·04	
15795·05 (2)	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 3→3, $Q3$	~$0{\cdot}14_1$	a	2	·088	0·015	?blend
			b	?1	4·947	0·05	
15793·89 (5)	α, 2→2, $P4$	—	z	1	·863	0·01	—
15787·23 (5)	α, 3→3, $Q2$ and	—	z	~15	·212	0·008	—
	?$^3A \rightarrow 2p\,^3\Pi$, 0→3, $Q2$ and do. ^3H (alt.)		?x	1	6·959	0·04	
15767·10 (7)	α, 3→3, $Q3$	—	z	1	·067	0·006	—
15726·90 (0)	$3p\,^1\Pi \rightarrow 2s\,^1\Sigma$, 1→0, $P3$	—	?x	—	·933	0·05	—
15726·30 (0)	unclassified	—			·216	0·05	
15723·84 (0)	unclassified	—	z	1	·816	0·01	—
15712·79 (6)	α, 2→2, $P5$ and	—	z	1	·781	0·006	—
	$3p\,^1\Pi \rightarrow 2s\,^1\Sigma$, 2→1, $Q1$						
15688·81 (5)	α, 3→3, $P2$	~0·05	z	8	·782	0·007	? asymmetric
			?x	1	—	—	not measured
15676·02 (1)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $R2$	—	z	?1	5·982	0·05	—
15669·39 (4)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $R1$	—	z	1	·351	0·007	—
15665·20 (0)	unclassified	—	z	?1	·207	0·03	—
15662·53 (2)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $R3$	—	z	1	·519	0·01	—
15650·27 (1)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 5→2, $R1$ and	—	z	~1	·252	0·01	—
	$^1N \rightarrow 2p\,^1\Pi$, 1→1, $Q1$						
15642·48 (1)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $R0$	—	z	1	·484	0·007	—
15629·10 (00)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $R4$	—	?z	—	·077	0·05	—
15626·94 (1)	α, 2→2, $P6$	—			·897	0·05	
15621·96 (10)	α, 3→3, $P3$	—	z	1	·960	0·005	—
15610·91 (1)	0→5, $^1\Pi_d \rightarrow 2p\,^1\Sigma$, $Q2$ and	—	x	1	~1·05	—	—
	?3→3, $3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, $Q5$	$0{\cdot}11_4$	a	1	·885	0·04	
			b (b)	? ~1/3	·771	0·04	
15607·91 (0)	unclassified	—	(b)	?1	·859	0·04	—
15591·98 (2)	unclassified	—	z	1	·952	0·007	—
15552·37 (10)	α, 4→4, $Q1$	—	z	1	·327	0·009	—
15539·88 (3)	α, 4→4, $Q2$	—	z	1	·313	0·01	—
15536·11 (2)	unclassified	0·155	a	2	·118	0·015	—
			b	1	5·963	0·015	
15528·92 (3)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $P1$	—	z	1	·885	0·007	—
15520·05 (3)	α, 4→4, $Q3$ and	—	z	?1	·001	0·007	seems sharp
	$3d\,^3\Sigma \rightarrow 2p\,^3\Pi$, 4→4, $Q1$						
15484·92 (1)	unclassified	—	z	1	·910	0·015	—
15448·06 (2)	$3p\,^3\Sigma \rightarrow 2s\,^3\Sigma$, 2→0, $P2$	—	z	~1	2·995	0·008	—

NOTES REFERRED TO IN TABLE 3, COLUMN 8

note number ν_G

1 17572·31 There may be a further component p_2 in which case the alternative structure is:

component	$I(a/b)$	ν_F	tolerance	notes
a	2·5	0·364	0·006	—
p_2	0·3	0·19	0·03	(unassigned)
b	1·0	0·14	0·02	—
p_1	0·5	0·10	0·02	(unassigned)

All b components are incompletely resolved; their mean $\nu_F = 17572{\cdot}127$.

2 17451·59 Pl. 86. Discharge tube pressure $\sim 0{\cdot}3_0$ mm. Hg current = 25 mA. Pl. 40. Discharge tube pressure $\sim 0{\cdot}16$ mm. Hg current = 15 mA. Lines considerably sharper with Pl. 40 conditions.

3 17441·32 $P2 = [17441{\cdot}21]$; this may be faint. There is no evidence for the bracketed intensity assignments except that $I(a/b) = \sim 4/1$ gives a mean ν which gives a good combination and is consistent with the 'regular and irregular' doublet scheme.

4 17433·50 Most of (x) might belong to 17436·24 (1 h).

5 17261·34 Gale *et al.* intensities for 17261·34 and 17260·69 apparently inverted by misprint.

6 17252·65 Line probably strengthens with increase of pressure and temperature. Bad combination with $R4$.

7 17218·50 $R2$ [17218·51], $Q'4$ [17218·47]. $I\ (R2/Q'4) = \sim [15K/8K]$. a and b have each been assigned to $R2$ and $Q'4$. $\Delta_c\ Q'4\ R'3$ interval seems satisfactory though this does not carry much weight because of poor measures on 17433·50. $I(a/b)\ R2$ should be $\sim 4/1$ and $I(a/b)\ Q'4 \sim 5/2$. These do not combine to give the observed result. On our plates the line is weak for (8) (G).

8 17144·93 Some of (a) probably belongs to 17143·50 (0) G.

9 17125·07 Both components probably confused: see 17004·40. There is very little confirmatory evidence of ours for ν_a by combinations. ? ?b component may be entirely due to 17127·12 (0) G.

10 17033·21 $I(Q3/Q4)$ probably >1. y assigned twice and it may also belong to 17031·67, 17030·88 and 17030·19.

11 17023·67 (q) and (b) unresolved. (?0) may be partly b of $1 \rightarrow 1$, $Q'2$.

12 17004·40 z (q) does not appear to belong to Π_d, $0 \rightarrow 0$, $P'2$; possibly x does. There may be other components masked by 17006·43.

13 16990·70 $1 \rightarrow 1$, $R3$ needs [3/4] of the total strength. Only 4/11 has been assigned. This analysis gives doublet separations larger than expected for $1 \rightarrow 1$ $R3$ and the two lines of 16988·72 q.v. with which the pattern is badly mixed.

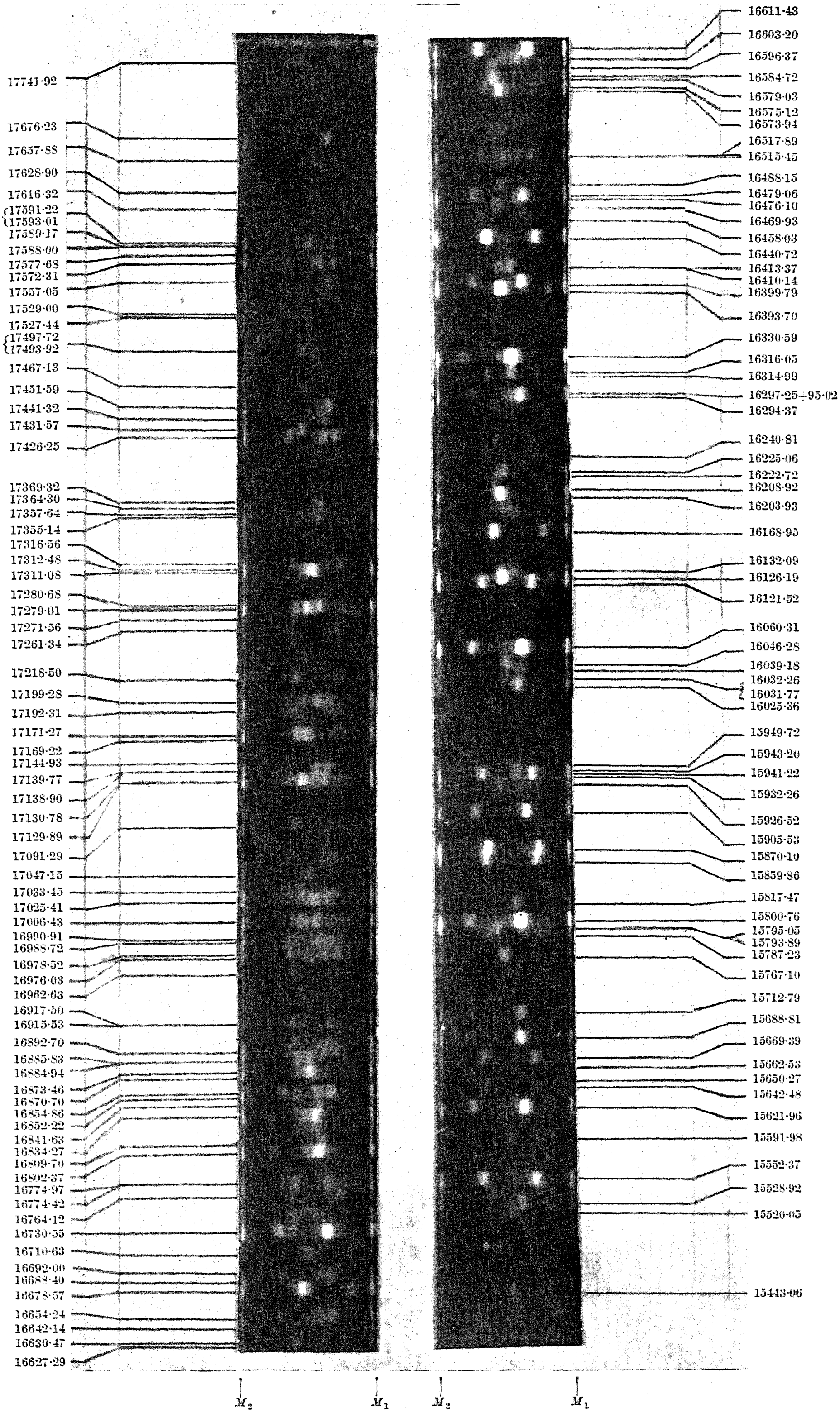
14 16988·72 See note 13. Δ_d, $1 \rightarrow 1$, $R'2$ needs [$\sim 1/3$] total strength. $Q1$ uncertain but expected to be $<2/3$ total strength. It is assumed that there are no overlapping orders in 88·72 and 90·71, otherwise there are insufficient data for a more detailed analysis.

15 16917·50, 16916·57, 16915·33 See § 11 for discussion of this group.

16 16841·63 ?p, or some of it, might be due to 16836·04 (1)

$$I(Q1:Q'2:Q3:Q'4) = [3K:1K:\leqslant 3K:\leqslant 0{\cdot}3K].$$

There appear to be no lines between x and y with $I > I(x) \div (4 \text{ or } 5)$ a possible interpretation is $y = Q1 + Q'2$, $x = Q3$, $?p = Q'4$ with intensities $I(Q1:Q'2:Q3:Q'4) \approx [7:2:3:\tfrac{1}{2}]$; but there are several other possibilities.



reflexion echelon spectrum of H_2 between 17741·92 and 15443·06 cm.$^{-1}$

(*Facing p.* 172)

note number	ν_G	
17	16834·27	(l), (m), (n) not clearly resolved.
18	16822·00	Very bad combination. Our results for the (*) lines of this group rather doubtful: there may be an alternative analysis.
19	16802·37	$I(a/b)$ seems to tend towards lower values with our higher temperature and pressure discharges.
20	16774·97	$P\,2$ [16775·06]: most of the strength in (a) component is due to the blend 16774·42. $\nu P\,2$ is probably a bit high. Only very feeble signs of a; (b), (b') may be feebly blended respectively with the a and b components of 16772·74 (1).
21	16764·12	Microphotometer gives $I(a/b) = 2{\cdot}9_5 \pm\sim 0{\cdot}3$.
22	16753·44	$I(a/b) \ngtr 4/2$ unless there is a blend (?x) with the low frequency components; there may be something in this belonging to 16752·69 (0a).
23	16596·37	Surprisingly faint for a G (10) line on our plates.
24	16517·89	The centre of intensity seems to shift towards the low ν end of this group with a high temperature source and most of the strength then seems to be in $a_1 b_1$ or $a_1 b_1$ and b_2, all of which are blended by the increased line width. This may account for the different ν's of different observers.
25	16488·15	(?a_3) and (a_2) are resolved on only one plate. There may be only one component and the apparent resolution a spurious effect of grain. Mean $\nu_F(a+a') = 16488{\cdot}17_5 \pm 0{\cdot}01$.
26	16469·93	Mean ν for a and b is $16469{\cdot}93_0 \pm 0{\cdot}005$ cm.$^{-1}$. ?x may be spurious.
27	?16458·03	ν may differ from the given value by n times (0·727) cm.$^{-1}$, where n is an integer (or zero), probably about 5.
28	16330·59	The z component probably contains b' of 16332·11 as a weak blend.
29	16039·18	The wave number may be variable (? blend). At the lowest pressures we get 16039·167 cm.$^{-1}$.
30	16032·09	Poor fit with Gale *et al.* Perhaps there are some overlapping orders giving, for example, a component at $\nu = \nu x - 0{\cdot}727$ cm.$^{-1}$.
31	16031·77	Remarks of note 30 apply except that $\nu = \nu z + 0{\cdot}727$ cm.$^{-1}$.

13. COMPARISON WITH RICHARDSON AND WILLIAMS' RESULTS FOR BANDS ENDING ON $2p\,^3\Pi$

This paper will end by comparing the splittings ($\Delta\nu$) and the intensity ratio $I(a/b)$ of the fine-structure components of lines of the bands ending on the $2p\,^3\Pi$ state as now found (table 3) with the values estimated by Richardson & Williams (1931). There are only six lines, all belonging to the system $3d\,^3\Sigma$–$2p\,^3\Pi$, and the data are set out in table 4.

TABLE 4

allocation and wave number		R. & W. (1931)		F. & R. (1944)		remarks
		$\Delta\nu$ about	$I(a/b)$ about	$\Delta\nu$	$I(a/b)$	
			0 → 0 *bands*			
$Q\,1$	16854·86 (10)	0·21	3	0·194	2·5	U means insufficiently resolved for a reliable estimate to be made
$Q\,3$	16654·24 (9)	0·17	U	0·173	2·5	
$R\,2$	16834·27 (10)	0·11	U	0·114	4, U	
$R\,4$	16710·63 (6)	U	U	0·06 U	*ca.* 1	
			1 → 1 *bands*			
$Q\,1$	16603·20 (9)	0·21	5/3	0·196	2·5	—
$R\,2$	16584·72 (10)	0·08	U	0·118	5	—

The values in table 4 represent complete agreement within the limits for which reliance could be placed, in the estimates made from the 1931 observations. Too much confidence should not be placed in the last decimal figure in the 1944 results whether for $\Delta\nu$ or $I(a/b)$, but some lines are better than others. Some guidance can be got from the figures in the tolerance column of table 3 with regard to this.

We wish to express our thanks to the Department of Scientific and Industrial Research and to the Royal Society for the provision of financial assistance which enabled this work to be carried out.

REFERENCES

Barrell, H. & Sears, J. E. 1939 *Phil. Trans.* A, **238**, 32.

Drinkwater, J. W., Richardson, Sir Owen & Williams, W. Ewart 1940 *Proc. Roy. Soc.* A, **174**, 164.

Gale, H. G., Monk, G. S. & Lee, K. O. 1928 *Astrophys. J.* **67**, 89.

Kapuscinski, W. & Eymers, J. G. 1929 *Proc. Roy. Soc.* A, **122**, 58.

Meggers, W. F. & Peters, C. G. 1918 *Bull U.S. Bur. Stand.* **14**, 697.

Merton, T. R. & Barratt, S. 1922 *Phil. Trans.* A, **222**, 369.

Richardson, O. W. 1934 *Molecular hydrogen and its spectrum.* New Haven: Yale University Press; London: Oxford University Press.

Richardson, O. W. & Williams, W. E. 1931 *Nature*, **127**, 481.

Schüler, H. 1931 *Z. Phys.* **72**, 423.

Williams, W. E. 1931 *Nature*, **127**, 816.

Williams, W. E. 1933 *Proc. Phys. Soc.* **45**, 699.

Williams, W. E. & Gogate, D. V. 1938 *Proc. Roy. Soc.* A, **167**, 509.

Williams, W. E. & Middleton, A. 1939 *Proc. Roy. Soc.* A, **172**, 159.

The fine structure of the lines of the H_2 spectrum which go down to the $2p\,^3\Pi$ states

By E. W. Foster and Sir Owen Richardson, F.R.S.

(*Received* 5 *December* 1944)

The echelon data tabulated and described in the preceding paper referred to here as Paper I are used to investigate the fine structure of the H_2 band lines discovered by Richardson & Williams in 1931. Practically all the lines with resolved fine structure have $2p\,^3\Pi$ as their lower states and the number of such resolved lines is now very greatly increased.

The new information is applied in the first instance to the critical examination and improvement of the existing tables of the (composite) $2p\,^3\Pi$ level differences. These are a basic tool in the elucidation of this spectrum. One result of this part of the investigation is that the pairs of lines from which each composite level difference is derived divide themselves into two groups which have slightly different values.

The fine structure shows itself visually as a resolution of the lines hitherto regarded as single into two components which we denote by *a* and *b*, *a* having the higher frequency and intensity. The differences for *a* or *b* are considered separately and shown to be different from each other and from the differences of the composite lines. A number of regularities among these new differences are pointed out.

The largest doublet separation $\Delta\nu$ observed is about 0·22 cm.$^{-1}$ and the smallest 0·06 cm.$^{-1}$. The lines (which go down from $3s\,^3\Sigma$ and $3d\,^3\Sigma$, Π, Δ to $2p\,^3\Pi_{cd}$) are divided into a 'regular' and an 'irregular' group. Lines of the regular group have a larger $\Delta\nu$ and the intensity ratio of *a* to *b* is nearly constant and close to 2·4, whereas for the irregular group this ratio falls from about 5 to about 1 as K'' increases from 1 to 4.

The regular group consists of all the lines which go down to $2p\,^3\Pi_c$, and the irregular group of all which go down to $2p\,^3\Pi_d$.

The greater part (about 90 %) of the doublet separations arises from fine structure splitting of the lower $2p\,^3\Pi$ levels. When effects arising from the upper levels are eliminated it is found that the splitting, both of $2p\,^3\Pi_c$ and $2p\,^3\Pi_d$, diminishes as K'' increases and at an increasing rate at higher K''.

In general these fine structures are closely parallel with those of He_2 discovered by Monk & Mulliken in 1929. The H_2 data cover a wider range of states and fill in the gaps left by the 'missing lines' in the He_2 spectrum. There are, however, some differences outstanding which seem to call for further investigation.

1. THE ROTATIONAL WAVE-NUMBER DIFFERENCES

The bands of the H_2 spectrum which have the lower triplet states $2p\,^3\Pi_a$ (or $2p\,^3\Pi_c$, an alternative notation) and $2p\,^3\Pi_b$ (or $2p\,^3\Pi_d$) were discovered by Richardson in 1929 (see Richardson 1929, 1930; Richardson & Davidson 1931; and Sandeman 1932). They are distinguished by the fact that in the really strong systems in which the transitions start from the complexes $3s\,^3\Sigma$ and 3, $4d\,^3\Sigma$, $^3\Pi_{cd}$, $^3\Delta_{cd}$ the intensity is almost entirely confined to the transitions of the axial vibrational sequence $0'$–$0''$, $1'$–$1''$, $2'$–$2''$, etc. For this reason progress in the analysis and classification of the lines of these bands has been governed almost entirely by an accurate knowledge of the rotational wave-number differences and particularly of those such as $Q\,1-P\,2 = R\,1-Q\,2$ which are determined solely by the final states $2p\,^3\Pi_c$ or $2p\,^3\Pi_d$.

This analysis is due to Richardson & Davidson (1931, 1933), Sandeman (1932), Richardson, Davidson, Marsden & Evans (1933) and to Richardson & Rymer (1934 *a*, *b*). It is based almost entirely on the wave-number and intensity tables of Gale, Monk & Lee (1928), the only ones which are both complete and accurate enough for the purpose. Now the present experimental work has introduced a new factor into this business. We find that all the structures in the bands ending on $2p\,^3\Pi$ which have hitherto been regarded as single lines are, in fact, multiples with at least two members. The separation of these 'doublets' is not constant, but in most cases the components are clearly resolved on our plates. Examples illustrating these statements can be recognized on plate 4 of Foster & Richardson entitled 'The fine structure of the spectral lines of H_2: measurements with a reflexion echelon', and lists of such examples are given in the footnotes on p. 159 of that paper which we shall refer to subsequently as Paper I. It follows that the wave-number differences, such as $Q\,1-P\,2$, etc., which we have been using, are not really numbers characteristic of the differences of certain energy levels of the states $2p\,^3\Pi$ but are differences of sums of the energies of at least two such states, weighted in proportions determined by intensity factors in a way which is irrelevant and disturbing to the elucidation of the energy-level differences which should be the really constant quantities.

The logical course to pursue in these circumstances would be to abandon the existing $Q\,1-P\,2$, etc., differences and to replace them by similar differences of corresponding components. Apart from the difficulty at this stage of being quite certain which are the corresponding components, it would seem that the results would be of no immediate use in furthering the elucidation of the H_2 spectrum, as even now all the available experimental data relate to what are very probably incompletely resolved lines. For these reasons we have postponed this task for the time being and adopted a course more likely to be of immediate service in the reduction of the spectrum.

The work of Richardson & Rymer (1934 *a*, *b*) shows that there are band systems ending on $2p\,^3\Pi_{cd}$ in the extreme violet as well as in the red and near-red part of the spectrum here dealt with, and it is likely that there are numerous band systems ending on $2p\,^3\Pi$ which are still undiscovered in various parts of the spectrum. Even in the near-red, some of the few still unclassified lines look like $2p\,^3\Pi$ 'doublets'. It is probable that the $Q\,1-P\,2$ and similar differences, got with resolution like that of Gale *et al.* in which the 'doublets' are not resolved, are constant enough for reducing data obtained with the same resolution, even if they are not strictly constants at all. Moreover, if they should ultimately not turn out to be constant but to show small systematic deviations of an ascertainable nature, that fact might well prove an important clue to the interpretation of the fine structure of the lines.

Having determined the wave numbers of the doublet constituents and measured the ratios of their intensities, the optical centre of the equivalent line with the doublets unresolved can be calculated.

However, we cannot strictly claim to have measured our intensity ratios directly. In a few instances the strengths of the *a* and *b* components were photographically the same. In these cases the equality of the blackening was checked, using a microphotometer (designed by W. H. J. Childs). Knowing the approximate position of the components relative to the position of the maximum of the intensity envelope and knowing the form of the envelope, the intensity ratio could be calculated. In the majority of cases the photographic intensities of *a* and *b* were not equal. In these cases the intensity ratio was estimated visually (with or without microphotometer traces) and a correction was applied for the envelope effect. In several such cases where photographic intensities were nearly the same the final corrected intensity ratios are probably fairly close to absolute values. In other cases the determinations suffer more from the shortcomings of visual intensity estimation. However, since the separation of all the doublets is of the same order the error in the wave number of the 'equivalent line' is largely offset (in these cases where the intensity ratio is comparatively large) by the comparatively small weight to be attached to the weaker component.

In view of all these considerations we decided to redetermine the 'constant differences' of the $Q\,1-P\,2$ type by building up the composite lines from our intensity and wave-number data for the resolved constituents. In doing this, we expected to get more accurate determinations not only on account of our more accurate measures but also because our superior resolution enabled us to eliminate the effects of numerous blends. In many cases their actual presence had been only a theoretical inference, not an observational fact, and the extent of the interference was often uncertain as the relative intensities of the blended lines were only roughly calculable.

The reason for introducing this examination and revision of the existing differences at the outset is not because we believe it is the most important result of the paper but because it is also a convenient, compact and critical way of assembling data necessary to establish other conclusions arrived at.

A knowledge of the notation used in connexion with the bands ending on $2p\,^3\Pi_{cd}$ will be assumed. It is given in Richardson & Davidson (1931, p. 658) and also in Richardson (1934, chapter VIII). The alternate weaker lines, a characteristic feature of all the bands which arise from transitions between the *s* levels, are distinguished by dashes, as $Q'\,1$, $P'\,2$, etc.

The data for the undashed lines, such as $R\,0$, $Q\,1$, $P\,2$, etc., are given in table 1 and those for the dashed lines in table 2. There are no combinations between dashed and undashed lines. The following notes will probably be helpful in interpreting the tables.

As already pointed out above, the first use we shall make of the material in tables 1 and 2 is to recompute the rotational differences of the $2p\,^3\Pi$ levels using the new echelon data. The results for the *a* terms are given in table 3 and those for the *s* terms in table 4.

It is difficult to make general statements about the accuracy of the values given in these tables. It is doubtful if it is as high as the agreement of many of the figures might suggest. In the first place, it often seems higher than the tolerance figures in tables 1 and 2 would lead us to anticipate. In the second place, individual agreements are often surprisingly good and in other cases unexpectedly poor. However, it seems certain that they are better in general than the old values got from the unresolved composite lines, and that it will be safe to substitute tables 3 and 4 for the corresponding parts of tables LXXVI and LXXVII of Richardson (1934). In doubtful cases, probably the best course would be to compromise by averaging the two values.

Fewer term differences are represented in table 3 and 4 than in the old tables. This is mainly because many of the weaker lines do not appear in the echelon patterns.

An examination of table 3 suggests quite definitely that the frequency differences $Qm-P(m+1)$ are not exactly equal to the differences $R(m)-Q(m+1)$. The differences between the mean values of the $Q(m)-P(m+1)$ and those of the $R(m)-Q(m+1)$ are larger than what one would expect to arise from the effect of the chance errors which operate within the individual groups. There is also evidence of a systematic change in the differences of the weighted means with K values in each band and with vibrational frequency on passing from one band to another. On the whole, the evidence for the reality of these differences is greatest where the observational data seem to be most satisfactory.

Apart from inferior resolution which might obscure them, there is no reason why these differences should not be found in the data in the older tables, such as Richardson (1934). The means $\overline{\delta\nu_R}$ from these tables are set alongside those of $\overline{\delta\nu_F}$ from tables 3 and 4 in table 5. The close resemblance between the differences of the $\overline{\delta\nu_F}$ and $\overline{\delta\nu_R}$ throughout table 5 can hardly be entirely accidental.

Having compiled these figures, it is of interest to see whether there is any evidence of systematic difference between $\overline{\delta\nu_F}$ and $\overline{\delta\nu_R}$. From the data in table 5 we have deduced the following information about the values of $\overline{\delta\nu_F}-\overline{\delta\nu_R}$: for the

$$Q(m)-P(m+1) \quad \text{and} \quad R(m)-Q(m+1)$$

differences there are fifteen negative values, total $-0{\cdot}425_5$ cm.$^{-1}$, average $-0{\cdot}0284$ cm.$^{-1}$, and seventeen positive values, total $+0{\cdot}6858$ cm.$^{-1}$, average $+0{\cdot}0403$ cm.$^{-1}$; the average of all thirty-two of these values is $+0{\cdot}00813$. For the $Q'm-P(m+1)$ and $R'(m)-Q'(m+1)$ differences there are nine negative values, total $-0{\cdot}314$, average $-0{\cdot}0349$, and ten positive values, total $+0{\cdot}231$, average $+0{\cdot}0231$, the average of all nineteen values being $-0{\cdot}00437$. The average of all the fifty-one values of $\overline{\delta\nu_F}-\overline{\delta\nu_R}$ is $+0{\cdot}00347$.

The only evidence of systematic variation in these figures is the positive average value for the undashed and the negative average value for the dashed line differences. This is probably caused by the fact that the undashed lines are on the average three times as intense as the dashed lines. The small positive residue for the whole fifty-one pairs of lines is about what would be expected from the larger number of undashed pairs.

If the apparent difference between the final terms of the lines coming from the Σ, Π_c, Δ_c levels on the one hand and the Π_d, Δ_d levels on the other is real, it should throw valuable light on the selection rules governing the fine-structure components of $2p\,^3\Pi$. If there were no fine structure such a difference would appear to be impossible, as has been assumed hitherto in the analysis of the $2p\,^3\Pi_{cd}$ levels. However, the observational evidence we have so far presented cannot be considered as conclusive, so we shall postpone the further consideration of this issue until more evidence bearing on it has accumulated.

2. THE FINAL ($Q(m)-P(m+1)$, ETC.) a AND b DIFFERENCES CONSIDERED SEPARATELY

So far we have been concerned principally with the unresolved lines as measured by previous observers or with the corresponding equivalents built up from our measures of the wave numbers of the resolved constituents and our estimates of their relative intensities. Our main object in this has been to eliminate the effects of overlapping lines which it has not hitherto been possible to allow for. This has been done largely with a view to increasing the reliability of existing tables as tools for future use.

When facing the ultimate interpretation of the data, one would expect to reach the goal more rapidly by operating with the a and b constituents separately rather than with the composite lines out of which they are built up. In any event, one can hardly hope to obtain a satisfactory comprehension of the whole without a good understanding of the operation of the parts. We have, therefore, extracted from tables 1 and 2 the differences $Q(m)-P(m+1)$ for the a (higher frequency, more intense) and b components respectively. These are set out in table 6 (undashed lines) and table 7 (dashed lines). Only the decimal parts of the wave-number differences are given; the integral parts are the same as those for the composite lines (tables 1 and 2).

In assessing the data in tables 6 and 7, it is to be remembered that they refer to pairs of lines which are not arbitrary, but the choice of one determines that of the other. This limitation greatly reduces both the number and accuracy of the data. A line measured very accurately may be quite useless because its partner is, for one of a variety of reasons, incapable of measurement. On the other hand, any 'bad' line destroys nearly all the virtue of a 'good' line with which it is associated. The reliability of any pair is little better than that of the less reliable component.

From examination of tables 6 and 7, the following conclusions may be drawn:

(1) The a differences are in general not the same as the b differences.

(2) The $Q(m)-P(m+1)$ differences are in general not the same as the

$$R(m)-Q(m+1)$$

differences, either for the a or b components or for the combination of them which makes up the composite line hitherto observed.

This statement about the combination also confirms the conclusion already drawn from the data for the unresolved lines.

(*continued on* p. 199)

TABLE 1. ACCURACY OF COMBINATIONS OF LINES

0→0 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p\,^3\Pi$	branch	K''	wave number ν_G	intensity G	intensity K	component	decimal of wave number ν_F	tolerance ±
$3d\,^3\Sigma$	Q	1	16854·86	10	37·5	a b	·870 ·676	0·004 0·005
	P	2	16734·20	4	11·0	a_2 ?b_2	$\cdot 15_5$ $\cdot 01_5$	0·04 0·05
$3d\,^3\Pi_d$	R	1	17199·28	10	44	a_2 b_2	·228 ·012	0·008 0·005
	P	3	16898·48	5	16·3	a_2 (B) b_2 (a)	·432 ·255	0·03 0·04
$3d\,^3\Pi_c$	Q	1	17212·01	1	—	?a ?b	·012 1·827	0·04 0·04
	P	2	17091·29	8	17·0	a b	·277 ·105	0·01 0·01
$3d\,^3\Delta_d$	R	1	17572·31	10	34·6	a b	·364 ·151	0·006 0·015
	Q	2	17451·59	10	30·7	a b	·593 ·444	— 0·006
$3d\,^3\Sigma$	R	2	16834·27	10	37·7	a (l) b (m)	·273 ·159	0·015 0·03
	Q	3	16654·24	9	18·1	a b	·253 ·080	0·006 0·006
$3d\,^3\Pi_c$	R	2	17441·32	8	19·6	a b	·347 ·178	0·006 0·01
	Q	3	17261·34	4	17·1	a b	·310 ·159	0·015 0·015
$3d\,^3\Delta_d$	Q	2	17451·59	10	30·7	a b	·593 ·444	— 0·006
	P	3	17271·56	4	13·9	a b	·552 ·408	0·01 0·02
$3d\,^3\Delta_c$	R	2	17717·06	3a	5·9	a b	abs. abs.	— —
	Q	3	17537·02	4	9·6	a b	7·023 6·807	0·03 0·03
$3d\,^3\Pi_d$	R	3	17199·28	10	44	a_1 b_1	·388 ·228	0·005 0·008
	Q	4	16961·69	0	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	Q	3	17261·34	4	17·1	a b	·310 ·159	0·015 0·015
	?P	4	17023·67	1	—	a_3 (?0) b_3 (?m)	·63 ·54	0·02 0·04
$3d\,^3\Delta_d$	R	3	17765·17	3	5	a b	abs. abs.	— —
	Q	4	17527·44	4	11·1	a b	U.	—
$3d\,^3\Delta_c$	Q	3	17537·02	4	9·6	a b	7·023 6·807	0·03 0·03
	P	4	17299·30	4	14·6	a b	U.	—
$3d\,^3\Sigma$	R	4	16710·63	6	21·8	a b	U.	0·006
	Q	5	16416·35	2	—	— —	abs. abs.	— —

ENDING ON $2p\,^3\Pi$. (*a*) 'STRONG LINES' (UNDASHED)

0→0 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
16854·814	0·194	6	5/2	3	120·66	120·68$_7$	120·741	obs.
16734·127	0·14$_0$	1	?2/1	0				4/1
17199·166	0·216	4	[5/2]	2	300·80	300·789	300·777	5/2
16898·382	0·177	3	5/2	2				5/2
17211·959	?0·185	1	?2/1	1	120·72	120·716	120·741	5/2
17091·243	0·172	3	4/1	3				obs.
17572·303	0·213	3	5/2	2	120·72	120·737	120·741	obs.
17451·566	0·149	3	?4·5/1	1				obs.
16834·250	0·114	3	4/1	3	180·03	180·046	180·036	obs.
16654·204	0·173	6	5/2	2				obs.
17441·313	0·169	2	5/2	0	179·98	180·046	180·036	4/1
17261·267	0·151	3	5/2	2				obs.
17451·566	0·149	3	?4·5/1	1	180·08	180·055	180·036	obs.
17271·511	0·144	3	5/2	1				obs.
$G\nu_F$17717·04	—	—	—	—	180·04	180·08	180·036	—
17536·961	0·216	2	5/2	1				obs.
17199·342	0·160	4	[5/2]	1	237·59	237·672	237·674	—
$G\nu_F$16961·67	—	—	—	—				—
17261·267	0·151	3	5/2	2	237·67	237·682	237·674	—
17023·585	0·09	1	?1/1	1				obs.
$G\nu_F$17765·15	—	—	—	—	237·73	237·737	237·674	—
17527·413	?0·08	1	?1·5/1	1				o.c.
17536·961	0·216	2	5/2	1	237·72	237·694	237·674	—
17299·267	0·08	3	?1/1·5	2				o.c.
16710·599	0·06	3	1/1	3	294·28	294·269	294·239	o.c.
$G\nu_F$16416·33	—	—	—	—				—

TABLE 1

0 → 0 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p\,^3\Pi$	branch	K''	wave number ν_G	intensity		component	decimal of wave number ν_F	tolerance ±
				G	K			
$3d\,^3\Pi_d$	Q	4	16961·69	0	—	— —	abs. abs.	— —
	P	5	16667·33	1	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	R	4	17547·00	2*a*	6·0	— —	— —	— —
	?Q	5	17252·65	2	11·0	*a* ?*b*	·639 ·49	0·03 0·05
$3d\,^3\Delta_d$	Q	4	17527·44	4	11·1	*a* *b*	U.	0·02
	P	5	17233·19	2*h*	10·3	— —	abs. abs.	— —
$3d\,^3\Delta_c$	R	4	17986·10	0	—	—	abs.	—
	Q	5	17691·96	0	—	—	abs.	—
$3d\,^3\Pi_d$	R	5	17179·50	3	9·1	—	abs.	—
	Q	6	16831·97	0	—	—	abs.	—
$3d\,^3\Delta_d$	R	5	17940·62	1	—	—	abs.	—
	Q	6	17593·01	3	3·1	—	abs.	—
$3d\,^3\Pi_d$	Q	6	16831·97	0	—	—	abs.	—
	P	7	16431·36	0	—	—	abs.	—
$3d\,^3\Delta_d$	Q	6	17593·01	3	3·1	—	abs.	—
	P	7	17192·31	6	22·9	—	abs.	—
1 → 1 BANDS								
$3s\,^3\Sigma$	Q	1	16917·50	3	10·9	a_1 (*A*) b_1 (*B*)	·539 ·346	0·02 0·02
	P	2	16802·37	10	52·0	*a* *b*	·361 ·211	0·015 0·02
$3d\,^3\Sigma$	Q	1	16603·20	9	27·0	*a* *b*	·247 ·051	0·006 0·008
	P	2	16488·15	8	30·6	a_1 b_1	8·056 7·95	0·03 0·04
$3d\,^3\Pi_d$	R	1	16978·52	10	28·0	*a* *b*	·548 ·343	0·005 0·006
	P	3	16692·00	5	14·0	*a* *b*	2·032 1·886	0·01 0·015
$3d\,^3\Pi_c$	Q	1	16988·72	8	22·7	a_2 (*A*) b_2 (*B*)	·635 ·415	0·02 0·02
	P	2	16873·46	9	39·4	*a* *b*	·453 ·271	0·015 0·01
$3d\,^3\Delta_d$	R	1	17426·25	10	40·3	*a* *b*	·276 ·086	0·004 0·006
	Q	2	17311·08	9	37·0	*a* ?*b*	1·078 ·940	0·02 —
$3s\,^3\Sigma$	R	2	17086·67	2	9·2	?*a* *b*	·70 ·475	0·06 0·02
	Q	3	16915·33	3	13·0	a_2 (*A*) b_2 (*B*)	·358 ·165	0·02 0·02
$3d\,^3\Sigma$	R	2	16584·72	10	34·7	*a* *b*	·729 ·611	0·01 0·02
	Q	3	16413·37	7	26·1	*a* *b*	·425 ·243	0·025 0·02

(*continued*)

0→0 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_B$	16 $\delta\nu_F$	17 $\delta\nu_B$ standard	18 intensity ratio (for calculation of mean ν_F)
$G\nu_F$16961·67	—	—	—	—	294·36	—	294·239	—
$G\nu_F$16667·31	—	—	—	—				—
$G\nu_F$17546·98	—	—	—	—	294·35	294·38	294·239	—
17252·60	?0·15	0	?2/1	1				5/2
17527·413	—	—	—	—	294·25	294·24	294·239	—
$G\nu_F$17233·17	—	—	—	—				—
—	—	—	—	—	294·14	—	294·239	— —
—	—	—	—	—	347·53	—	347·60	— —
—	—	—	—	—	347·61	—	347·60	— —
—	—	—	—	—	400·61	—	?400·63	— —
—	—	—	—	—	400·70	—	?400·63	— —
1→1 BANDS								
16917·484	0·20*	1	?5/2	1	115·13	115·144	115·140	obs.
16802·340	0·150	3	4–6/1	4				6/1
16603·191	0·196	5	5/2	2	115·05	115·161	115·140	obs.
16488·030	0·106	2	3/1	2				obs.
16978·489	0·205	6	5/2	2	286·52	286·499	286·479	obs.
16691·990	0·146	4	5/2	1				obs.
16988·572	0·220	2	?2/1	0	115·26	115·149	115·140	5/2
16873·423	0·182	3	5/1	3				obs.
17426·222	0·190	6	1·9/1	5	115·17	115·163	115·140	5/2
17311·059	0·138	2	?6/1	1				obs.
17086·662	?0·225	0	?3/1	0	171·34	171·359	171·339	5/1
16915·303	0·18*	1	?5/2	1				obs.
16584·709	0·118	3	5/1	4	171·35	171·336	171·339	obs.
16413·373	0·182	2	5/2	2				obs.

TABLE 1

1 → 1 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p\,^3\Pi$	branch	K''	wave number ν_G	intensity		component	decimal of wave number ν_F	tolerance ±
				G	K			
$3d\,^3\Pi_c$	R	2	17218·50	8	23·0	a b	·494 ·376	0·015 0·03
	Q	3	17047·15	4	15·0	a b	·189 ·019	0·005 0·01
$3d\,^3\Delta_d$	Q	2	17311·08	9	37·0	a ?b	1·078 0·940	0·02 —
	P	3	17139·77	5	17·6	a (x) b (y)	771 ·599	0·01 0·03
$3d\,^3\Delta_c$	R	2	17529·00	5	12·0	a ?b	8·992 abs.	0·02 —
	Q	3	17357·64	6	15·6	a b	·693 ·493	0·02 0·02
$3s\,^3\Sigma$	Q	3	16915·33	3	13·0	a_2 (A) b_2 (B)	·358 ·165	0 02 0·02
	P	4	16688·40	4	19·0	— (p) —	·373 —	0·02 —
$3d\,^3\Pi_d$	R	3	16990·71	6	21·3	a_1 (b) b_1 (?x)	·711 ·50	0·05 0·05
	P	5	16484·00	1	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	Q	3	17047·15	4	15·0	a b	·189 ·019	0·005 0·01
	P	4	16820·29	2	14·0	$\left.\begin{matrix}a_4\\b_4\end{matrix}\right\}$ (p)	? broad	0·06
$3d\,^3\Delta_d$	R	3	17591·22	4	7·1	a_1 (b) b_1	·222 ? ·07	0·03 0·06
	Q	4	17364·30	5	11·7	a b	·350 ·216	0·02 0·02
$3d\,^3\Delta_c$	Q	3	17357·64	6	15·6	a b	·693 ·493	0 02 0·02
	P	4	17130·78	3a	—	a_1 b_1	[·78] [·66]	— —
$3d\,^3\Sigma$	R	4	16469·93	8	22·6	a b	·997 ·897	0·02 0·02
	Q	5	16190·17	3	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	R	4	17335·25	2	8·9	—	abs.	—
	Q	5	17055·50	1	—	—	abs.	—
$3d\,^3\Delta_d$	Q	4	17364·30	5	11·7	a b	·350 ·216	0·02 0·02
	P	5	17084·56	2	9·8	?a b	·56 ·48	0·06 0·03
$3d\,^3\Delta_c$	R	4	17739·54	1	—	—	abs.	—
	Q	5	17459·76	2	—	—	abs.	—

2 → 2 BANDS

1	2	3	4	5	6	7	8	9
$3d\,^3\Sigma$	Q	1	16297·25	5	22·3	a (r) b (q)	·294 ·137	0·03 0·05
	P	2	16187·50	1	—	a_2 (?y) b_2 (a)	·556 ·383	0·03 0·04
$3d\,^3\Pi_d$	R	1	16730·55	10	12·0	a b	·709 ·520	0·01 0·04
	P	3	16458·03	5	20·8	a b	8·040 7·886	0·004 0·006

(*continued*)

1 → 1 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
17218·455	0·118	2	2/1	0	171·35	171·315	171·339	obs.
17047·140	0·170	4	5/2	2				obs.
17311·059	?0·138	2	?6/1	1	171·31	171·337	171·339	obs.
17139·722	0·172	2	?5/2	0				5/2
17528·967	[0·15]	0	>4/1	0	171·36	171·331	171·339	5/1
17357·636	0·200	2	5/2	1				obs.
16915·303	0·18*	1	?5/2	1	226·93	226·930	226·909	—
16688·373	—	—	—	—				—
16990·658	?0·21	0	3/1	0	506·71	506·676	506·679	obs.
$G\nu_F$16483·975	—	—	—	—				—
17047·140	0·170	4	5/2	2	226·86	226·90	226·909	—
16820·24	?0·12	1	?1/1	0				o.c.
17591·184	?0·15$_2$	1	?3/1	0	226·92	226·901	226·909	obs.
17364·283	0·134	3	?1/1	0				obs.
17357·636	0·200	2	5/2	1	226·86	—	226·909	—
[17130·727] $G\nu_F$17130·76	[0·12]	—	—	—				—
16469·942	0·10$_0$	3	4/5	5	279·76	279·792	279·770	obs.
$G\nu_F$16910·15	—	—	—	—				—
—	—	—	—	—	279·75	—	279·770	—
17364·283	0·134	3	?1/1	0	279·74	279·76$_3$	279·770	obs.
17084·52	?0·08	1	?1/1	0				obs.
—	—	—	—	—	279·78	—	279·770	—

2 → 2 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
16297·249	0·157	1	?3/1	0	109·75	109·728	109·770	5/2
16187·521	0·173	1	?4/1	0				obs.
16730·655	0·189	3	?5/2	0	272·52	272·659	272·59$_0$	5/2
16457·996	0·154	5	5/2	2				obs.

TABLE 1

2 → 2 BANDS

1 transition to $2p\,^3\Pi$	2 branch	3 K''	4 wave number ν_G	5 intensity G	6 K	7 component	8 decimal of wave number ν_F	9 tolerance ±
$3d\,^3\Pi_c$	Q	1	16684·88	1	—	?a (x) ?b (p)	·912 ·738	0·01 0·04
	P	2	16575·12	5	24·5	a b	5·137 4·97	0·02 0·05
$3d\,^3\Delta_d$	R	1	17279·01	7	46·4	a b	9·015 8·830	0·007 0·004
	Q	2	17169·22	10	37·8	a b	·232 ·087	0·01 0·03
$3d\,^3\Sigma$	R	2	16295·02	3	—	?a_1 (q) b_1 (p)	5·03 4·87	0·05 0·05
	Q	3	16132·09	5	18·4	a b	2·134 1·968	0·01 0·015
$3d\,^3\Pi_c$	R	2	16897·19	5	10·9	a_3 (A) b_3	·226 abs.	0·02 —
	Q	3	16734·20	4	11·0	?a_1 b_1 (r)	·325 ·155	0·05 0·04
$3d\,^3\Delta_c$	R	2	17355·14	6	16·2	a b	·17 ·04	0·05 0·04
	Q	3	17192·31	6	22·9	a b	·280 ·130	0·006 0·01
$3d\,^3\Delta_d$	Q	2	17169·22	10	37·8	a b	·232 ·087	0·01 0·03
	P	3	17006·43	10	30·9	a_1 b_1	·350 ·184	0·02 0·02
$3d\,^3\Pi_c$	Q	3	16734·20	4	11·0	?a_1 b_1 (r)	·325 ·155	0·05 0·04
	P	4	16517·89	6	16·8	a (b_2) b (a_1)	8·024 7·906	0·02 0·02
$3d\,^3\Delta_c$	Q	3	17192·31	6	22·9	a b	·280 ·130	0·006 0·01
	P	4	16976·03	5	17·5	$\left.\begin{matrix}a\\b\end{matrix}\right\}$ (x)	? broad	0·02
$3d\,^3\Pi_d$	R	3	16753·44	4	15·0	a b	·450 ·296	0·005 0·015
	P	5	16271·37	2	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	R	4	17013·50	1	—	—	abs.	—
	Q	5	?16747·70	1	—	—	abs.	—
$3d\,^3\Delta_d$	R	3	17416·51	3	—	—	abs.	—
	Q	4	17200·19	1	—	—	abs.	—
$3d\,^3\Sigma$	R	4	16192·25	6	15·8	a_1 b_1	·312 ·191	0·03 0·01
	Q	5	15926·52	3h	—	$\left.\begin{matrix}a\\b\end{matrix}\right\}$ (?p)	? broad	—
$3d\,^3\Delta_d$	Q	4	17200·19	—	—	— —	abs. abs.	— —
	P	5	16934·41	1	9·8	?a (y) ?b (x)	·463 ·323	0·04 0·04
$3d\,^3\Pi_c$	Q	5	?16747·70	1	—	—	abs.	—
	P	6	?16431·36	0	—	—	abs.	—
$3d\,^3\Delta_d$	R	5	17541·45	2	—	—	abs.	—
	Q	6	17224·94	1	—	—	abs.	—
$3d\,^3\Delta_c$	Q	5	?17233·19	2h	10·3	—	abs.	—
	P	6	16916·57	0	—	—	abs.	—
$3d\,^3\Pi_d$	R	5	16746·17	1	—	—	abs.	—
	P	7	16068·21	2	—	—	abs.	—

(*continued*)

2→2 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
16684·862	?0·174	1	?3/1	0	109·76	109·758	109·770	5/2
16575·104	0·167	1	4/1	2				obs.
17278·962	0·185	5	5/2	2	109·79	109·754	109·770	obs.
17169·208	0·145	2	5/1	1				obs.
16294·99$_8$	0·160	1	?4/1	0	162·93	162·911	162·8$_0$	4/1
16132·087	0·166	3	5/2	2				obs.
16897·196	[0·15]	0	≥4/1	0	162·99	162·919	162·8$_9$	4/1
16734·277	0·170	1	[5/2]	0				5/2
17355·14$_4$	0·13	1	4/1	1	162·83	162·907	162·8$_0$	obs.
17192·237	0·150	4	5/2	2				obs.
17169·208	0·145	2	5/1	1	162·79	162·906	162·8$_0$	—
17006·302	0·166	3	?5/2	1				—
16734·277	0·170	1	[5/2]	0	216·31	216·316	216·30$_3$	5/2
16517·960	0·118	3	5/6	2				obs.
17192·237	0·150	4	5/2	2	216·28	216·305	216·30$_3$	obs.
16975·932	—	—	≥3/5	0				o.c.
16753·388	0·154	3	3/2	2	482·07	482·03$_8$	482·06$_8$	3/2
$G\nu_F$16271·35	—	—	—	—				—
—	—	—	—	—	—	—	265·76$_5$	—
—	—	—	—	—	—	—	216·30$_3$	—
16192·264	0·121	2	3/2	1	265·73	265·785	265·76$_5$	obs.
?15926·479	—	—	—	—				—
$G\nu_F$17200·17	—	—	—	—	265·78	265·75	265·765	—
16934·423	0·140	1	?2/1	1				5/2
—	—	—	—	—	316·34	—	316·5$_5$	—
—	—	—	—	—	316·51	—	316·5$_5$	—
—	—	—	—	—	316·62	—	316·5$_5$	—
	—	—	—	—	677·96	—	—	—

TABLE 1

3→3 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p^3\Pi$	branch	K''	wave number ν_G	intensity G	intensity K	component	decimal of wave number ν_F	tolerance ±
$3d^3\Sigma$	*Q*	1	15931·94	2	—	*a*	2·015	0·008
						b	1·849	0·02
	P	2	15827·48	1	—	*a*	·495	0·03
						b	abs.	—
$3d^3\Pi_d$	*R*	1	16445·80	3	14·6	*a*	·823	0·05
						b	·699	0·04
	P	3	16186·63	3	15·5	a_3 (*a*)	·656	0·03
						b_3 (*b*)	·524	0·03
$3d^3\Pi_c$	*Q*	1	16388·32	1	—	*a*	·366	0·02
						?*b*	·155	0·03
	P	2	16283·83	2	8·6	—	u.d.	—
						—	u.d.	—
$3d^3\Delta_d$	*R*	1	17129·89	8	45·2	*a*	·917	0·008
						b	·735	0·005
	Q	2	17025·41	6	25·6	a_2	·405	0·05
						b_2 (*p*)	·239	0·08
$3d^3\Sigma$	*R*	2	15949·72	3	—	*a*	·767	0·03
						b	·632	0·03
	Q	3	15795·05	2	—	*a*	5·088	0·015
						b	4·947	0·05
$3d^3\Pi_d$	*Q*	2	16341·29	0	—	—	abs.	—
						—	abs.	—
	P	3	16186·63	3	15·5	a_3 (*a*)	·656	0·03
						b_3 (*b*)	·524	0·03
$3d^3\Pi_c$	*R*	2	16579·03	2	10·8	—	u.d.	—
						—	u.d.	—
	Q	3	16424·30	00	—	—	abs.	—
						—	abs.	—
$3d^3\Delta_d$	*Q*	2	17025·41	6	25·6	a_2	·405	0·05
						b_2 (*p*)	·239	0·08
	P	3	16870·70	2	10·5	*a*	·742	0·015
						b (*b′*)	·591	0·02
$3d^3\Delta_c$	*R*	2	17187·90	4	13·7	—	u.d.	—
						—	u.d.	—
	Q	3	17033·21	4	20·2	*a* (*z*)	·217	0·02
						b (*y*)	·057	0·04
$3d^3\Pi_d$	*R*	3	16476·10	7	32·3	*a* (*x*)	·226	0·03
						b (*p*)	·057	0·05
	P	5	16018·16	1	—	—	abs.	—
						—	abs.	—
$3d^3\Pi_c$	*Q*	3	16424·30	00	—	—	abs.	—
	P	4	16218·29	—	—	—	abs.	—
$3d^3\Delta_d$	*R*	3	17239·70	2	9·9	—	abs.	—
						—	abs.	—
	Q	4	[17033·66]	—	—	*a* (*y*)	·783	0·05
						b (*x*)	·616	0·04
$3d^3\Delta_c$	*Q*	3	17033·21	4	20·2	*a*	·217	0·02
						b	·057	0·04
	P	4	16827·14	2	7·9	—	u.d.	—
						—	u.d.	—
$3d^3\Sigma$	*R*	4	15862·86	1*h*	—	—	abs.	—
						—	abs.	—
	Q	5	?15610·91	1	—	*a*	·885	0·04
						b	·771	0·04
$3d^3\Delta_d$	*Q*	4	[17033·66]	—	—	*a* (*y*)	·783	0·05
						b (*x*)	·616	0·04
	P	5	16781·72	1*a*	—	—	abs.	—
						—	abs.	—
$3d^3\Delta_c$	*R*	4	?17282·98	1	—	—	abs.	—
						—	abs.	—
	Q	5	?17030·88	1	—	—	abs.	—
						—	abs.	—

(*continued*)

3 → 3 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
15931·968	0·166	3	5/2	1	104·46	104·503	$104{\cdot}48_2$	obs.
15827·465	[0·15]	0	[4/1]	0				4/1
16445·787	0·124	1	5/2	1	259·17	259·164	$259{\cdot}18_2$	obs.
16186·623	0·132	2	?3/1	1				3/1
16388·306	0·211	2	5/2	0	104·49	104·496	104·482	obs.
16283·81	—	—	—	—				—
17129·865	0·182	6	5/2	2	104·48	104·488	104·482	obs.
17025·377	$0{\cdot}16_6$	1	[?5/1]	0				5/1
15949·713	0·135	2	3/2	2	154·67	154·665	$154{\cdot}70_0$	obs.
15795·048	0·141	2	?2/1	1				5/2
$G\nu_F$16341·27	—	—	—	—	154·66	154·647	$154{\cdot}70_0$	—
16186·623	0·132	2	?3/1	1				3/1
—	—	—	—	—	154·73	—	$154{\cdot}70_0$	—
—	—	—	—	—				—
17025·377	$0{\cdot}16_6$	1	[?5/1]	0	154·71	154·678	$154{\cdot}70_0$	5/1
16870·699	0·151	3	2/1	1				5/2
$G\nu_F$17187·88	—	—	—	—	154·69	154·71	$154{\cdot}70_0$	—
17033·171	0·160	2	[5/2]	0				—
$16476{\cdot}17_8$	0·169	1	[5/2]	0	457·94	$458{\cdot}03_8$	458·03	5/2
$G\nu_F$16018·14	—	—	—	—				—
16424·28	—	—	—	—	206·01	—	206·04	—
16218·27	—	—	—	—				
17239·68	—	—	—	—	[206·04]	205·98	206·04	—
17033·699	0·167	1	?1/1	0				obs.
17033·171	0·160	2	[5/2]	0	206·07	$206{\cdot}05_1$	206·04	—
$G\nu_F$16827·12	—	—	—	—				—
$G\nu_F$15862·84	—	—	—	—	251·95	$251{\cdot}98_8$	251·990	—
$15610{\cdot}85_2$	0·114	1	?3/1	1				5/2
17033·699	0·167	1	?1/1	0	251·94	252·00	251·990	obs.
$G\nu_F$16781·70	—	—	—	—				—
?17282·96	—	—	—	—	252·10	—	251·990	—
$G\nu_F$?17030·86	—	—	—	—				—

TABLE 2. ACCURACY OF COMBINATIONS OF LINE

0→0 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p^3\Pi$	branch	K''	wave number ν_G	intensity G	intensity K	component	decimal of wave number ν_F	tolerance ±
$3d^3\Sigma$	R'	1	16884·94	4	22·3	a (y) ?b	·896 ? ·75	0·01 0·04
	Q'	2	16764·12	4	12·0	a b	4·148 3·959	0·008 0·005
$3d^3\Pi_d$	Q'	1	17125·07	2	—	a ??b	5·035 $4{\cdot}83_5$	0·05 0·05
	P'	2	17004·40	3a	14·9	— —	— —	— —
$3d^3\Pi_c$	R'	1	17369·32	5	—	a b	·307 abs.	0·006 —
	Q'	2	17248·55	1	—	a b	·522 abs.	0·02 —
$3d^3\Delta_c$	R'	1	17588·00	5	22·0	a_2 (d) b_2 (c)	7·962 7·833	0·02 0·04
	Q'	2	17467·13	6	10·0	a b	·220 060	0·007 0·02
$3d^3\Pi_c$	Q'	2	17248·55	1	—	a b	·522 abs.	0·02 —
	P'	3	17068·60	1	—	$\left.\begin{matrix}a\\b\end{matrix}\right\}$	U.	—
$3d^3\Delta_d$	R'	2	17669·81	2	3·3	— —	abs. abs.	— —
	Q'	3	17489·90	3	9·6	— —	abs. abs.	— —
$3d^3\Delta_c$	Q'	2	17467·13	6	10·0	a b	·220 ·060	0·007 0·02
	P'	3	17287·27	2	9·3	?a, z b	·252 abs.	0·03 —
$3d^3\Sigma$	R'	3	16774·97	4	U.	a_1 (b) b_1	·870 ·75	0·02 0·05
	Q'	4	16536·89	1	8·7	a b	— —	— —
$3d^3\Pi_c$	R'	3	17498·65	2 U.	—	—	abs.	—
	Q'	4	17260·69	0	—	—	abs.	—
$3d^3\Delta_d$	Q'	3	17489·90	3	9·6	—	abs.	—
	P'	4	17251·99	0a	—	—	abs.	—
$3d^3\Delta_d$	R'	4	17855·86	0	—	—	abs.	—
	Q'	5	17562·23	3	—	—	abs.	—
$3d^3\Delta_c$	Q'	4	17612·63	1	—	—	abs.	—
	P'	5	17318·92	1	—	—	abs.	—
$3d^3\Sigma$	R'	5	16643·08	1	—	—	abs.	—
	Q'	6	16295·02	3	—	—	abs.	—
$3d^3\Pi_c$	R'	5	17581·42?	1	—	—	abs.	—
	Q'	6	17233·19?	2h	10·3	—	abs.	—
$3d^3\Delta_d$	Q'	5	17562·23	3	—	—	abs.	—
	P'	6	17213·82	0	—	—	abs.	—

ENDING ON $2p\,^3\Pi$. (*s*) ‘WEAK LINES’ (DASHED)

0 → 0 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
16884·869	0·146	1	?4·5/1	0	120·82	120·775	120·79	obs.
16764·094	0·189	5	2·95/1	6				5/2
17125·002	??0·20	1	?5/1	0	120·67	120·62	120·79	obs.
$G\nu_F$17004·38 [17004·21]	—	—	—	—				—
17369·282	[0·15]	—	> 5/1	0	120·77	120·812	120·79	5/1
17248·470	[0·18]	0	[5/2]	0				5/2
17587·936	0·129	2	4/1	1	120·87	120·762	120·79	5/1
17467·174	0·160	3	2·5/1	1				5/2
17248·470	[0·18]	0	[5/2]	0	179·95	179·913	179·905	5/2
17068·557	?0·10	1	?1/1	0				—
—	—	—	—	—	179·91	—	179·905	—
—	—	—	—	—				—
17467·174	0·160	3	5/2	2	179·86	179·962	179·905	5/2
17287·212	[0·16]	0	—	—				3/1
16774·85	0·12	1	[4/1]	0	238·08	237·98	237·93	4/1
16536·87	—	—	—	—				—
—	—	—	—	—	237·96	—	237·93	—
—	—	—	—	—	237·91	—	237·93	—
—	—	—	—	—	293·63	—	293·72	—
—	—	—	—	—	293·71	—	293·72	—
—	—	—	—	—	348·06	—	348·42	—
—	—	—	—	—	348·23	—	348·42	—
	—	—	—	—	348·41	—	348·42	—

TABLE 2

1→1 BANDS

1	2	3	4	5	6	7	8	9
				intensity				
transition to $2p\,^3\Pi$	branch	K''	wave number ν_G	G	K	component	decimal of wave number ν_F	tolerance ±
$3s\,^3\Sigma$	R'	1	17031·67	1a	—	— —	abs. abs.	— —
	Q'	2	16916·57	0	—	?a (B) b	·628 ·466	0·04 0·02
$3d\,^3\Sigma$	R'	1	16630·47	3	8·4	a b	·468 abs.	0·015 —
	Q'	2	16515·45	5	15·0	a b	·488 ·309	0·006 0·008
$3d\,^3\Pi_d$	Q'	1	?16899·76	0	—	a_1 (b) b_1	·53_1 abs.	0·04 —
	P'	2	?16784·56	1	—	— —	abs. abs.	— —
$3d\,^3\Pi_c$	R'	1	17138·90	4h	—	a (y) b	·872 abs.	0·02 —
	Q'	2	?17023·67	1	—	a_4 b_4	[·86] [·70]	— —
$3d\,^3\Delta_c$	R'	1	17431·57	6	17·1	a_2 b_2	·559 abs.	0·006 —
	Q'	2	17316·56	5	15·6	a b	·579 ·418	0·008 0·015
$3s\,^3\Sigma$	Q'	2	16916·57	0	—	?a (B) b	·628 ·466	0·04 0·02
	P'	3	16744·92	5	16·7	— —	abs. abs.	— —
$3d\,^3\Pi_c$	R'	1	17138·90	4h	—	a (y) b	·872 abs.	0·02 —
	P'	3	16852·22	3	10·2	a ??b	·174 ·008	0·005 0·03
$3d\,^3\Delta_d$	R'	2	17508·99	5	12·1	—	abs.	—
	Q'	3	17337·39	3	11·5	—	abs.	—
$3d\,^3\Delta_c$	Q'	2	17316·56	5	15·6	a b	·579 ·418	0·008 0·015
	P'	3	17144·93	4	10·2	a ?b	·853 ·75	0·02 0·05
$3s\,^3\Sigma$	R'	3	17139·77	5	17·6	—	abs.	—
	Q'	4	16913·48	0	—	—	abs.	—
$3d\,^3\Pi_d$	R'	2	16988·72	8	22·7	a_1 (y) b_1 (A)	·864 ·635	0·03 0·04
	P'	4	16590·80	1	8·0	— —	abs. abs.	— —
$3d\,^3\Sigma$	R'	3	16530·23	5	14·2	a ?b (p)	·234 ·11	0·03 0·04
	Q'	4	16303·84	1	—	— —	— —	— —
$3d\,^3\Pi_c$	R'	3	17282·98	1	—	—	abs.	—
	Q'	4	17056·70	1	9·3	—	abs.	—
$3d\,^3\Delta_d$	Q'	3	17337·39	3	11·5	—	abs.	—
	P'	4	17111·01	2	—	—	abs.	—
$3d\,^3\Delta_c$	R'	3	17633·88	1	—	—	abs.	—
	Q'	4	17407·51	1	—	—	abs.	—
$3s\,^3\Sigma$	Q'	4	16913·48	0	—	—	abs.	—
	P'	5	16633·03	0	—	—	abs.	—
$3d\,^3\Delta_d$	R'	4	17669·81	2	3·3	—	abs.	—
	Q'	5	17389·47	2	—	—	abs.	—
$3d\,^3\Delta_c$	Q'	4	17407·51	1	—	—	abs.	—
	P'	5	17127·12	0	—	—	abs.	—
$3d\,^3\Sigma$	R'	5	16405·53	0a	—	—	abs.	—
	Q'	6	16074·39	0	—	—	abs.	—

(*continued*)

1 → 1 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\delta\nu_R$	16 $\delta\nu_F$	17 $\delta\nu_R$ standard	18 intensity ratio (for calculation of mean ν_F)
$G\nu_F$17031·65	—	—	—	—	115·10	115·06	115·03	—
16916·587	?0·162	1	?3/1	0				obs.
16630·443	[0·15]	0	⩾5/1	2	115·02	115·006	115·03	5/1
16515·437	0·179	5	[5/2]	1				obs.
16899·506	[0·15]	0	—	—	115·20	114·966	115·03	5/1
$G\nu_F$16784·54	—	—	—	—				—
17138·847	[0·15]	0	[5/1]	0	115·23	—	115·03	5/1
$G\nu_F$17023·65 [17023·817]	[0·16]	0	[5/2]	0				—
17431·534	[0·15]	0	[5/1]	0	115·01	115·001	115·03	5/1
17316·533	0·161	4	5/2	2				obs.
16916·587	?0·162	1	?3/1	0	171·65	$171{\cdot}68_7$	171·65	—
$G\nu_F$16744·90	—	—	—	—				—
17138·847	—	—	—	—	286·68	286·710	286·68	—
16852·137	0·166	3	3·5/1	2				obs.
—	—	—	—	—	171·60	—	171·65	—
—	—	—	—	—				
17316·533	0·161	4	5/2	2	171·63	171·701	171·65	—
17144·832	$0{\cdot}01_3$	2	4/1	2				—
$G\nu_F$17139·75	—	—	—	—	226·29	—	$226{\cdot}34_2$	—
$G\nu_F$16913·46	—	—	—	—				
16988·80	0·229	0	[5/2]	0	397·92	398·02	$397{\cdot}99_2$	5/2
$G\nu_F$16590·78	—	—	—	—				—
16530·208	0·12	1	3·5/1	1	226·39	$226{\cdot}38_8$	226·342	3·5/1
$G\nu_F$16303·82	—	—	—	—				—
—	—	—	—	—	226·28	—	226·342	—
—	—	—	—	—				
—	—	—	—	—	226·38	—	226·342	—
—	—	—	—	—				
—	—	—	—	—	226·37	—	226·342	—
—	—	—	—	—				
—	—	—	—	—	280·45	—	280·400	-
—	—	—	—	—				
—	—	—	—	—	280·34	—	280·400	—
—	—	—	—	—				
—	—	—	—	—	280·39	—	280·400	—
—	—	—	—	—				
—	—	—	—	—	331·14	—	331·20	—
—	—	—	—	—				

TABLE 2

2→2 BANDS

1	2	3	4	5	6	7	8	9
				intensity			decimal of wave number	
transition to $2p\,^3\Pi$	branch	K''	wave number G.M.L.	G	K	component	ν_F	tolerance ±
$3d\,^3\Delta_d$	Q'	5	17389·47	2	—	—	abs.	—
	P'	6	17058·20	0	—	—	abs.	—
$3d\,^3\Sigma$	R'	1	16332·11	1	—	a_1	·141	0·04
						b_1	abs.	—
	Q'	2	16222·72	2	10·0	a	·743	$0{\cdot}02_5$
						b	·575	$0{\cdot}01_5$
$3d\,^3\Pi_d$	Q'	1	16648·40	2	—	a	·411	0·02
						b	abs.	—
	P'	2	16538·98	2	7·4	a	·99	0·04
						b	·88	0·06
$3d\,^3\Delta_c$	R'	1	17280·68	3	17·5	a	·661	0·02
						?b	·565	0·05
	Q'	2	17171·27	4	13·9	a	·290	0·03
						b	·115	0·01
$3d\,^3\Pi_c$	R'	1	16822·00	2	6·7	?a_3 (r)	1·95	0·08
						?b_3 (q)	1·78	0·08
	P'	3	16548·90	3	11·1	—	abs.	—
$3d\,^3\Delta_d$	R'	2	17348·12	3	9·4	a	·144	0·02
						?b	·006	0·03
	Q'	3	17184·46	3	10·7	a	·428	0·05
						b	·320	0·05
$3d\,^3\Delta_c$	Q'	2	17171·27	4	13·9	a	·290	0·03
						b	·115	0·01
	P'	3	17007·59	1	—	—	abs.	—
$3d\,^3\Sigma$	R'	3	17247·26	2	—	—	abs.	—
	Q'	4	16032·09	1a	—	$\left.\begin{array}{l}?a\\?b\end{array}\right\}\ \left.(x)\right\}$	·134	—
$3d\,^3\Pi_c$	R'	3	16961·25	0h	—	—	abs.	—
	Q'	4	16746·17	1	—	—	abs.	—
$3d\,^3\Delta_d$	Q'	3	17184·46	3	10·7	a	·428	0·05
						b	·320	0·05
	P'	4	16969·39	1	—	—	abs.	—
$3d\,^3\Delta_c$	R'	3	17433·50	1	—	??a_1 (?y)	·590	0·05
						?b_1 (x)	·378	0·02
	Q'	4	17218·50	8	23·0	a	·494	$0{\cdot}01_5$
						b	·376	0·03
$3d\,^3\Delta_d$	R'	4	17481·61	1	—	—	abs.	—
	Q'	5	17214·05	U.	—	—	abs.	—
$3d\,^3\Delta_c$	Q'	4	17218·50	8	23·0	a	·494	$0{\cdot}01_5$
						b	·376	0·03
	P'	5	16950·98	0a	—	—	—	—
$3d\,^3\Delta_d$	Q'	5	17214·05	U.	—	—	abs.	—
	P'	6	16899·76	0	—	—	abs.	—
$3d\,^3\Sigma$	P'	1	16827·69	1	—	—	abs.	—
$3d\,^3\Pi_d$	R'	2	16746·56	2	—	—	abs.	—
$3d\,^3\Pi_d$	R'	4	16752·69	0a	—	—	abs.	—
$3d\,^3\Pi_d$	P'	4	16367·90	1	8·4	—	abs.	—

(*continued*)

2→2 BANDS

10 mean wave number $\bar{\nu}_F$	11 $\Delta\nu$ (splitting)	12 weight of $\Delta\nu$ figure	13 intensity ratio (observed) $I(a/b)$	14 weight of $I(a/b)$ figure	15 $\Delta'\nu$ G.M.L.	16 $\Delta'\nu_F$	17 $\Delta'\nu$ G.M.L. standard	18 intensity ratio (for calculation of mean ν)
—	—	—	—	—	331·27	—	331·20	—
—	—	—	—	—				
16332·116	[0·15]	0	—	0	109·39	109·421	$109{\cdot}40_0$	5/1
16222·695	0·168	3	5/2	2				obs.
16648·386	[0·15]	0	⩾4/1	0	109·42	109·426	$109{\cdot}40_0$	5/1
16538·96	0·11	0	3/1	1				—
17280·647	0·096	2	?6/1	5	109·41	109·407	$109{\cdot}40_0$	obs.
17171·246	0·175	3	3/1	2				5/2
16821·92	?0·17	1	?5/2	0	273·10	273·04	$273{\cdot}09_2$	4/1
$G\nu_F$16548·88	—	—	—	—				—
17348·105	0·138	2	?5/2	1	163·66	163·704	163·692	obs.
17184·401	?0·108	1	?3/1	0				3/1
17171·246	—	—	—	—	163·68	163·676	$163{\cdot}69_2$	—
$G\nu_F$17007·57	—	—	—	—				—
—	—	—	—	—	215·00	$215{\cdot}10_6$	$215{\cdot}0_7$	—
16032·134	—	—	—	—				—
—	—	—	—	—	215·08	—	$215{\cdot}0_7$	—
—	—	—	—	—				
17184·401	—	—	—	—	215·07	$215{\cdot}03_1$	$215{\cdot}0_7$	—
$G\nu_F$16969·37	—	—	—	—				—
17433·537	$?0{\cdot}21_2$	0	?3/1	0	215·00	215·077	$215{\cdot}0_7$	obs.
17218·460	0·118	1	2/1	0				5/2
—	—	—	—	—	267·56	—	$267{\cdot}5_2$	
—	—	—	—	—				
17218·460	—	—	—	—	267·52	$267{\cdot}50_0$	$267{\cdot}5_2$	—
$G\nu_F$16950·96	—	—	—	—				—
—	—	—	—	—	314·29	—	$314{\cdot}_3$	—
—	—	—	—	—				
—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—

TABLE 2

3→3 BANDS

1	2	3	4	5	6	7	8	9
transition to $2p^3\Pi$	branch	K''	wave number ν_G	intensity		component	decimal of wave number ν_F	tolerance ±
				G	K			
$3d^3\Sigma$	R'	1	?15976·14	0	—	— —	abs. abs.	— —
	Q'	2	?15872·20	1	—	a b (?x)	[·271] ·101	— 0·02
$3d^3\Pi_d$	Q'	1	16363·49	1	—	—	abs.	—
	P'	2	16259·53	1	—	—	abs.	—
$3d^3\Pi_c$	R'	1	16511·46	1	—	a ?b	·470 ·32	0·03 0·05
	P'	3	16251·79	4	12·0	a_1 ?b_1	·820 ·679	0·02 0·05
$3d^3\Delta_c$	R'	1	[17130·30?]	—	—	a b	·276 ·10	0·02 0·06
	Q'	2	17026·36?	2	—	a_1 (a) b_1 (b)	·368 ·189	0·02 0·06
$3d^3\Delta_d$	R'	2	17186·00	2	8·9	?a ?b	?6·018 ?5·882	0·08 0·08
	Q'	3	17030·19	1	—	— —	abs. abs.	— —
$3d^3\Delta_c$	Q'	2	17026·36?	2	—	a_1 (a) b_1 (b)	·368 ·189	0·02 0·06
	P'	3	16870·70? (0·36 M.B.)	2	10·5	a_1 (b') b_1	·591 abs.	0·02 —
$3d^3\Sigma$	R'	3	15910·77	0	—	—	abs.	—
	Q'	4	15706·57	1	—	—	abs.	—
$3d^3\Pi_c$	R'	3	16633·03	0	—	—	abs.	—
	Q'	4	16428·77	0	—	—	abs.	—
$3d^3\Delta_d$	Q'	3	17030·19	1	—	— —	abs. abs.	— —
	P'	4	16825·98	1	—	?a_2 (x) ??b_2	6·027 5·917	0·03 0·05
$3d^3\Delta_c$	R'	3	17244·91	0	—	— —	abs. abs.	— —
	Q'	4	17040·75	1	—	a b	·803 ·675	0·02 0·03
$3d^3\Pi_d$	R'	2	16465·55	2	—	?a ?b	·654 ·480	0·04 0·02
	P'	4	16105·59	1	—	— —	abs. abs.	— —
$3d^3\Delta_d$	R'	4	17287·27	2	9·3	?a (z)	·252	0·03
	Q'	5	?17036·21	ab	—	—	abs.	—
$3d^3\Sigma$	P'	1	15913·37	00	—	—	abs.	—

Notes on tables 1 *and* 2

1. The figures given for 'tolerances' in column 9 are those given in column 7, table, 3, Paper I; the method of deriving them is explained in the general notes to table 3, Paper I.

2. The figures of column 12 for weighting the $\Delta\nu$'s should be a simple function of the corresponding figures of column 9, but in actual fact they were not compiled from those figures but were based on a second review of the data and thus form an independent set of estimates, and these figures have been chosen for the final weighting for computing mean splittings of the different levels (table 3).

(*continued*)

3→3 BANDS

10	11	12	13	14	15	16	17	18
mean wave number $\bar{\nu}_F$	$\Delta\nu$ (splitting)	weight of $\Delta\nu$ figure	intensity ratio (observed) $I(a/b)$	weight of $I(a/b)$ figure	$\delta\nu_R$	$\delta\nu_F$	$\delta\nu_R$ standard	intensity ratio (for calculation of mean ν_F)
$G\nu_F$15976·12	—	—	—	—	103·94	103·90	103·94$_4$	—
?15872·222	[0·17]	0	—	0				5/2
— —	— —	— —	— —	— —	103·96	—	103·94$_4$	—
16511·433	0·15	1	?3/1	0	259·67	259·641	259·759	obs.
16251·792	0·141	1	4/1	0				obs.
17130·247	0·176	1	5/1	0	103·94	103·930	103·944	obs.
17026·317	0·179	1	?5/2	0				obs.
17185·98	??0·136	1	??5/2	0	155·81	155·81	155·8$_{15}$	5/2
$G\nu_F$17030·17	—	—	—	—				—
17026·317	0·179	1	?5/2	0	155·66	155·77	155·8$_{15}$	obs.
16870·55	[0·15]	0	—	0				3/1
— —	— —	— —	— —	— —	204·20	—	204·22$_8$	—
— —	— —	— —	— —	— —	204·26	—	204·22$_8$	—
$G\nu_F$17030·17	—	—	—	—	204·21	204·17$_5$	204·22$_8$	—
16825·995	?0·11	1	⩾4/1	0				5/2
$G\nu_F$17244·89	—	—	—	—	204·16	204·12	204·22$_8$	—
17040·766	?0·128	2	?2/1	0				5/2
16465·604	0·174	1	?1/1	1	359·96	360·03	360·043	5/2
$G\nu_F$16105·57	—	—	—	—				—
17287·252 $G\nu_F$17036·19	— —	— —	— —	— —	251·06	251·06	?254·06	—
—	—	—	—	—	—	—	—	—

3. When figures are given in square brackets, either they are entirely predictions or are predictions with only very feeble direct experimental evidence.

4. The wave numbers ν_G in column 4 are from the tables of Gale *et al.* (1928) and so are the eye estimates of intensity in column 5. The intensities under K (column 6) are the energy measures of Kapuscinsky & Eymers (1929).

5. The figures (column 8) under decimal of wave number ν_F are our measures for the fine-structure components a, b (column 7) of the unresolved composite lines measured by Gale *et al.* (column 4). The doublet splittings $\Delta\nu_F$ in column 11 are the

differences of the a and b values of ν_F in column 8. The weights 0–6 in column 12 divide the $\Delta\nu_F$'s of column 11 into seven groups such that we should expect those with the same number to have about the same accuracy, those with the highest accuracy having the highest number (6).

6. Column 13 gives the ratio of the intensity of the component a in column 8 to that of b. These are mostly eye estimates, but some were based on photometric measures. We believe that this relative accuracy is fairly assessed by the series of numbers 0–6 in column 14 in the same way as those in column 6 represent the accuracy of the splittings.

7. The mean wave numbers $\bar{\nu}_F$ in column 10 are the values which we calculate for the wave numbers of the composite lines from our measurements of the ν_F's of the components in column 8 using the intensity ratio $I(a/b)$ given in column 18. These are not always the same as our observed ratios given in column 14. Some of the observed ratios are known to be wrong owing to various causes. We have made an attempt to eliminate the effects of these in the values given in column 18 which differ from those given in column 13. Unfortunately, it is often difficult to find out to what extent we have been successful in this.

8. The values $\delta\nu_R$ in column 15 are the wave-number differences

$$F(m+1)-F(n) = R(m)-Q(m-1) = Q(m)-P(m-1)$$

of consecutive rotational terms found by Richardson & Davidson (1933). In most cases, these are based entirely on the measures of Gale *et al.* and are just the differences of successive pairs of ν_G's in column 4. The values $\delta\nu_F$ in column 16 are the same differences when the wave-lengths of the composite lines have been constructed from our measured wave numbers of the constituent doublets and are the differences of the successive pairs of values in column 10.

The 'standard' $\delta\nu_R$ values in column 17 are the weighted means of all available $\delta\nu_R$ values for the same pairs of final rotational levels published in Richardson & Davidson (1933).

9. When the letters u.d. occur in these tables, they replace one of the following expressions in the table in its original form: 'unreliable data', 'not measured', 'probably double', 'no useful data', 'not observed', 'nothing due to this line observed', 'too weak to have visible effect on pattern', 'not measured, probably double', 'only one component observed, too faint to be sure of detail'. These phrases have different meanings, but the main point for the purpose of this paper is the effect of what they express on our measured or estimated numbers, and this has been taken into account in estimating the weights and tolerances in the relevant columns.

10. The letter 'U.' in these tables means 'unresolved' and 'o.c.' means 'setting on optical centre'.

11. In a few cases tolerances are not quoted in column 9. In these cases reference should be made to table 3, Paper I.

12. For discussion of the asterisked $\Delta\nu$'s of column 11, table 1, 1→1 bands, see discussion in Paper I, p. 159.

(*End of notes on tables* 1 *and* 2)

(*continued from* p. 179)

(3) There is only one case when values of $R'(m)-Q'(m+1)$ and $Q'(m)-P'(m+1)$ are available for the same m (for $m=1$ in the 2-2 bands), but in this case the corresponding statement to (2) for the dashed transitions is applicable.

(4) If by $Q(m)-P(m+1)\,a$ and $Qm-P(m+1)\,b$ one denotes the $Q(m)-P(m+1)$ a and b differences and so on, then

$$\begin{array}{ll} Q1-P2a>Q1-P2b, & R1-Q2a>R1-Q2b, \\ Q2-P3a<Q2-P3b, & R2-Q3a<R2-Q3b, \\ Q3-P4a>Q3-P4b, & R3-Q4a>R3-Q4b, \end{array}$$

There are three cases in the 2–2 and 3–3 bands which show small differences on the wrong side of these inequalities, but they are all doubtful; otherwise, they apply without exception to the results for all the bands, 0–0, 1–1, 2–2 and 3–3, for which data are available.

(5) For the dashed transitions the data are much less numerous and also less representative. There are seven examples of $R'1-Q'2$ a and b for all of which $R'1-Q'2a$ is $<R'1-Q'2b$. The rest consist of one example each of $Q'1-P'2$ for which $Q'1-P'2a$ is $>Q'1-P'2b$, of $R'2-Q'3$, for which $R'2-Q'3a$ is $>R'2-Q'3b$ of $Q'2-P'3$, for which $Q'2-P'3a$ is $>Q'2-P'3b$ and of $R'3-Q'4$ for which $R'3-Q'4a$ is $>R'3-Q'4b$. Thus in nine out of the eleven examples the inequalities are in the opposite sense to those for the corresponding undashed transitions. It looks, therefore, as if there is in the main a reversal of sign of these inequalities in the dashed as compared with the undashed transitions. The two exceptions both occur in the 2–2 bands. There were three rather dubious exceptions to the general statement in the undashed transitions, all in the 2–2 and 3–3 bands. It may be that the rule only applies strictly to the lower vibrational levels and is tending to break down at the higher levels. The evidence is too scanty to say more than this.

(6) There is also evidence of another regularity when we compare the

$$R,\,Q(m)-Q,\,P(m+1)$$

a and b differences which come from the $3d\,{}^3\Pi_d,\ \Delta_d\rightarrow 2p\,{}^3\Pi$ bands with the

$$Q,\,P(m)-R,\,Q(m+1)$$

a and b differences which come from $3d\,{}^3\Sigma,\ \Pi_c,\ \Delta_c\rightarrow 2p\,{}^3\Pi$. Taking the mean values given in table 6 we have:

0→0 *bands* $\overline{R1-Q2}\,a-\overline{Q1-P2}\,a=+0{\cdot}038$, $\overline{R1-Q2}\,b+\overline{Q1-P2}\,b=+0{\cdot}027$

$\overline{Q2-P3}\,a-\overline{R2-Q3}\,a=+0{\cdot}008$, $\overline{Q2-P3}\,b-\overline{R2-Q3}\,b=-0{\cdot}012$

1→1 *bands* $\overline{R1-Q2}\,a-\overline{Q1-P2}\,a=+0{\cdot}017$, $\overline{R1-Q2}\,b-\overline{Q1-P2}\,b=+0{\cdot}007$

$\overline{Q2-P3}\,a-\overline{R2-Q3}\,a=+0{\cdot}009$, $\overline{Q2-P3}\,b-\overline{R2-Q3}\,b=-0{\cdot}036$

2→2 *bands* $\overline{R1-Q2}\,a-\overline{Q1-P2}\,a=+0{\cdot}037$, $\overline{R1-Q2}\,b-\overline{Q1-P2}\,b=-0{\cdot}021$

$\overline{Q2-P3}\,a-\overline{R2-Q3}\,a=+0{\cdot}008$, $\overline{Q2-P3}\,b-\overline{R2-Q3}\,b=-0{\cdot}009$

3→3 *bands* $\overline{R1-Q2}\,a-\overline{Q1-P2}\,a=-0{\cdot}012$

$\overline{Q2-P3}\,a-\overline{R2-Q3}\,a=-0{\cdot}018$, $\overline{Q2-P3}\,b-\overline{R2-Q3}\,b=-0{\cdot}029$

TABLE 3. FINAL ANTISYMMETRIC (*a*) ROTATIONAL TERMS

0 → 0 BANDS

initial state	$Q1-P2$	$R1-Q2$	$R2-Q3$	$Q2-P3$	$Q3-P4$	$R3-Q4$	$R4-Q5$	$Q4-P5$
$3d^3\Sigma$	$120{\cdot}68_7$	—	180·046	—	—	—	$294{\cdot}27_9$	—
$3d^3\Pi_d$	—	$120{\cdot}730_5$*	—	180·052†	—	$237{\cdot}68_2$	—	—
$3d^3\Pi_c$	120·716	—	180·046	—	237·682	—	294·37	—
$3d^3\Delta_d$	—	120·737	—	180·055	—	$237{\cdot}72_7$	—	$294{\cdot}25_3$
$3d\Delta_c$	—	—	180·07	—	237·694	—	—	—
weighted means	120·716	120·734	180·046	$180{\cdot}053_5$	237·688	$237{\cdot}70_5$	$294{\cdot}3_{25}$	$294{\cdot}2_{53}$
differences	−0·018		$-0{\cdot}007_5$		$-0{\cdot}01_7$		$+0{\cdot}0_{72}$	

* Got by subtracting $3d^3\Delta_d$, $R1-Q2 = 120{\cdot}737$ from $300{\cdot}784 = R1-P3$ of $3d^3\Pi_d \rightarrow 2p^3\Pi$.
† Got by subtracting $3d^3\Delta_d$, $Q2-P3 = 180{\cdot}053_5$ from $300{\cdot}784 = R1-P3$ of $3d^3\Pi_d \rightarrow 2p^3\Pi$.

1 → 1 BANDS

initial state	$Q1-P2$	$R1-Q2$	$R2-Q3$	$Q2-P3$	$Q3-P4$	$R3-Q4$	$R4-Q5$	$Q4-P5$
$3s^3\Sigma$	115·144	—	171·359	—	226·930	—	—	—
$3d^3\Sigma$	115·161	—	171·336	—	—	—	$279{\cdot}78_2$	—
$3d^3\Pi_d$	—	115·164*	—	171·335†	—	226·911‡	—	$279{\cdot}77_0$§
$3d^3\Pi_c$	115·149	—	171·315	—	226·90	—	—	—
$3d^3\Delta_d$	—	115·163	—	171·337	—	226·901	—	$279{\cdot}76_3$
$3d^3\Delta_c$	—	—	171·331	—	—	—	—	—
weighted means	115·151	$115{\cdot}163_5$	171·335	171·336	$226{\cdot}91_5$	226·906	$279{\cdot}78_2$	$279{\cdot}76_6$
differences	$-0{\cdot}012_5$		−0·001		$+0{\cdot}00_9$		$+0{\cdot}01_6$	

* Got by subtracting $Q2-P3$ from $3d^3\Pi_d$, $R1-P3 = 286{\cdot}499$.
† Got by subtracting $R1-Q2$ from $3d^3\Pi_d$, $R1-P3 = 286{\cdot}499$.
‡ Got by subtracting $Q4-P5$ from $3d^3\Pi_d$, $R3-P5 = 506{\cdot}676$.
§ Got by subtracting $R3-Q4$ from $3d^3\Pi_d$, $R3-P5 = 506{\cdot}676$.

2 → 2 BANDS

initial state	$Q1-P2$	$R1-Q2$	$R2-Q3$	$Q2-P3$	$Q3-P4$	$R3-Q4$	$R4-Q5$	$Q4-P5$
$3d^3\Sigma$	109·728	—	162·911	—	—	—	265·785	—
$3d^3\Pi_d$	—	109·753*	—	162·905†	—	$216{\cdot}30_8$‡	—	—
$3d^3\Pi_c$	109·758	—	162·919	—	216·316	—	—	—
$3d^3\Delta_d$	—	109·754	—	162·919	—	—	—	265·74
$3d^3\Delta_c$	—	—	162·907	—	216·305	—	—	—
weighted means	$109{\cdot}74_3$	$109{\cdot}753_5$	$162{\cdot}911_3$	$162{\cdot}905_5$	$216{\cdot}310_5$	$216{\cdot}30_8$	265·785	265·74
differences	$-0{\cdot}01_{05}$		$+0{\cdot}005_8$		$+0{\cdot}00_{25}$		$+0{\cdot}04_5$	

* Got by subtracting $Q2-P3$ from $3d^3\Pi_d$, $R1-P3 = 272{\cdot}659$.
† Got by subtracting $R1-Q2$ from $3d^3\Pi_d$, $R1-P3 = 272{\cdot}659$.
‡ Got by subtracting $Q4-P5$ from $3d^3\Pi_d$, $R3-P5 = 482{\cdot}048$.

3 → 3 BANDS

initial state	$Q1-P2$	$R1-Q2$	$R2-Q3$	$Q2-P3$	$Q3-P4$	$R3-Q4$	$R4-Q5$	$Q4-P5$
$3d^3\Sigma$	104·503	—	154·665	—	—	—	$251{\cdot}97_8$	—
$3d^3\Pi_d$	—	104·486*	—	$154{\cdot}64_7$ 154·677†	—	$206{\cdot}03_9$‡	—	$252{\cdot}07_7$§
$3d^3\Pi_c$	$104{\cdot}50_6$	—	—	—	—	—	—	—
$3d^3\Delta_d$	—	104·488	—	154·678	—	$205{\cdot}97_1$	—	$252{\cdot}00_9$
$3d^3\Delta_c$	—	—	$154{\cdot}69_9$	—	$206{\cdot}06_1$	—	—	—
weighted means	104·504	104·487	$154{\cdot}68_2$	154·678	$206{\cdot}06_1$	$206{\cdot}00_5$	$251{\cdot}97_8$	$252{\cdot}04_3$
differences	+0·017		$+0{\cdot}00_4$		$+0{\cdot}05_6$		$-0{\cdot}06_5$	

* Got by subtracting $Q2-P3$ from $3d^3\Pi_d$, $R1-P3 = 259{\cdot}164$.
† Got by subtracting $R1-Q2$ from $3d^3\Pi_d$, $R1-P3 = 259{\cdot}164$.
‡ Got by subtracting $3d^3\Delta_d$, $Q4-P5$ from $3d^3\Pi_d$, $R3-P5 = 458{\cdot}04_8$.
§ Got by subtracting $3d^3\Delta_d$, $R3-Q4$ from $3d^3\Pi_d$, $R3-P5 = 458{\cdot}04_8$.

TABLE 4. FINAL SYMMETRIC (*s*) ROTATIONAL TERMS

0→0 BANDS

initial state	$Q'1-P'2$	$R'1-Q'2$	$R'2-Q'3$	$Q'2-P'3$	$Q'3-P'4$	$R'3-Q'4$	$R'4-Q'5$	$Q'4-P'5$
$3d^3\Sigma$	—	120·775	—	—	—	237·98	—	—
$3d^3\Pi_d$	$120{\cdot}6_3$	—	—	—	—	—	—	—
$3d^3\Pi_c$	—	120·812	—	179·913	—	—	—	—
$3d^3\Delta_d$	—	—	—	—	—	—	—	—
$3d^3\Delta_c$	—	120·762	—	179·962	—	—	—	—
weighted means	$120{\cdot}6_3$	$120{\cdot}78_3$	—	$179{\cdot}93_7$	—	$237{\cdot}9_8$	—	—
differences	$-0{\cdot}1_5$		—		—		—	

1→1 BANDS

initial state	$Q'1-P'2$	$R'1-Q'2$	$R'2-Q'3$	$Q'2-P'3$	$Q'3-P'4$	$R'3-Q'4$	$R'4-Q'5$	$Q'4-P'5$
$3s^3\Sigma$	—	115·05	—	$171{\cdot}69_7$	—	—	—	—
$3d^3\Sigma$	—	$115{\cdot}00_6$	—	—	—	—	—	—
$3d^3\Pi_d$	114·98	—	—	—	—	—	—	—
$3d^3\Pi_c$	—	—	—	171·706*	—	—	—	—
$3d^3\Delta_d$	—	—	—	—	—	—	—	—
$3d^3\Delta_c$	—	115·001	—	171·701	—	—	—	—
weighted means	114·98	115·004	—	171·703	—	—	—	—
differences	$-0{\cdot}02_4$		—		—		—	

* From $3d^3\Pi_c$, $R'1-P'3$ and $R'1-Q'2$.

2→2 BANDS

initial state	$Q'1-P'2$	$R'1-Q'2$	$R'2-Q'3$	$Q'2-P'3$	$Q'3-P'4$	$R'3-Q'4$	$R'4-Q'5$	$Q'4-P'5$
$3d^3\Sigma$	—	109·42	—	163·64†	—	215·096	—	—
$3d^3\Pi_d$	$109{\cdot}42_6$	—	—	—	—	—	—	—
$3d^3\Pi_c$	—	—	—	—	—	—	—	—
$3d^3\Delta_d$	—	—	163·704	—	$215{\cdot}04_1$	—	—	—
$3d^3\Delta_c$	—	109·407	—	$163{\cdot}68_6$	—	215·077	—	—
weighted means	$109{\cdot}42_6$	109·407	163·704	$163{\cdot}66_5$	$215{\cdot}04_1$	215·077	—	—
differences	$+0{\cdot}01_9$		$+0{\cdot}03_9$		$-0{\cdot}03_6$		—	

† From $3d^3\Pi_c$, $R'1-P'3$ and $3d^3\Delta_c$, $R'1-Q'2$.

3→3 BANDS

initial state	$Q'1-P'2$	$R'1-Q'2$	$R'2-Q'3$	$Q'2-P'3$	$Q'3-P'4$	$R'3-Q'4$	$R'4-Q'5$	$Q'4-P'5$
$3d^3\Sigma$	—	103·91	—	—	—	—	—	—
$3d^3\Pi_d$	—	—	155·88‡	—	$204{\cdot}22_5$§	—	—	—
$3d^3\Pi_c$	—	—	—	—	—	—	—	—
$3d^3\Delta_d$	—	—	155·82	—	$204{\cdot}16_5$	—	$251{\cdot}07_2$	—
$3d^3\Delta_c$	—	103·930	—	155·77	—	$204{\cdot}11_4$	—	—
weighted means	—	103·930	155·85	155·77	$204{\cdot}19_5$	$204{\cdot}11_4$	$251{\cdot}07_2$	—
differences	—		$+0{\cdot}08$		$+0{\cdot}08_1$		—	

‡ From $3d^3\Pi_d$, $R'2-P'4$ and $3d^3\Delta_d$, $Q'3-P'4$.
§ From $3d^3\Pi_d$, $R'2-P'4$ and $3d^3\Delta_d$, $R'2-Q'3$.

TABLE 5. COMPARISON OF $\delta\nu_F$'S AND $\delta\nu_R$'S.
(THE NEW AND OLD FINAL ROTATIONAL TERMS)

	$Q1-P2$	$R1-Q2$	$R2-Q3$	$Q2-P3$	$Q3-P4$	$R3-Q4$	$R4-Q5$	$Q4-P5$
$\nu'' = 0$								
mean $\delta\nu_F$	120·716	120·734	180·046	180·053$_5$	237·688	237·70$_5$	294·3$_{25}$	294·2$_{53}$
differences	−0·018		−0·007$_5$		−0·01$_7$		+0·0$_{72}$	
mean $\delta\nu_R$	120·732	120·750	180·035	180·037	237·695	237·653	294·240	294·237
differences	−0·018		−0·002		+0·042		+0·003	
$\nu'' = 1$								
mean $\delta\nu_F$	115·151	115·163$_5$	171·335	171·336	226·91$_5$	226·906	279·78$_2$	279·766
differences	0·012$_5$		−0·001		+0·00$_9$		+0·01$_6$	
mean $\delta\nu_R$	115·110	115·140	171·338	171·340	226·91$_3$	226·90$_5$	279·76$_8$	279·84$_5$
differences	−0·030		−0·002		+0·00$_8$		−0·07$_7$	
$\nu'' = 2$								
mean $\delta\nu_F$	109·74$_3$	109·753$_5$	162·911$_3$	162·905$_5$	216·310$_5$	216·30$_8$	265·785	265·74
differences	−0·01$_{05}$		+0·005$_8$		+0·00$_{25}$		+0·04$_5$	
mean $\delta\nu_R$	109·775	109·76$_2$	162·85	162·76	216·29$_5$	216·3$_2$	265·76$_3$	265·7$_8$
differences	+0·01$_3$		+0·09		−0·0$_{25}$		−0·0$_{17}$	
$\nu'' = 3$								
mean $\delta\nu_F$	104·504	104·487	154·68$_2$	154·678	206·06$_1$	206·00$_5$	251·97$_8$	252·04$_3$
differences	+0·017		+0·00$_4$		+0·05$_6$		−0·06$_5$	
mean $\delta\nu_R$	104·473	104·493	154·690	154·710	206·040	206·$_{14}$	252·005	251·975
differences	−0·020		−0·020		+0·080		+0·03$_0$	

	$Q'1-P'2$	$R'1-Q'2$	$R'2-Q'3$	$Q'2-P'3$	$Q'3-P'4$	$R'3-Q'4$	$R'4-Q'5$	$Q'4-P'5$
$\nu'' = 0$								
mean $\delta\nu_F$	120·6$_3$	120·783	—	179·93$_7$	—	237·98	—	—
differences	−0·1$_{53}$		—		—		—	
mean $\delta\nu_R$	120·73	120·80	179·905	179·905	237·91	237·96	293·75	293·69
differences	−0·07		0·000		−0·05		+0·06	
$\nu'' = 1$								
mean $\delta\nu_F$	114·98	115·004	—	171·703	—	—	—	—
differences	−0·024		—		—		—	
mean $\delta\nu_R$	—	115·03	171·6$_0$	171·66$_3$	226·3$_8$	226·33$_3$	280·3$_4$	280·42$_0$
differences	—		−0·0$_{63}$		+0·0$_{47}$		−0·0$_{80}$	
$\nu'' = 2$								
mean $\delta\nu_F$	109·42$_6$	109·407	163·704	163·66$_5$	215·04$_1$	215·077	267·56	267·51$_0$
differences	+0·01$_9$		+0·03$_9$		−0·03$_6$		+0·05$_0$	
mean $\delta\nu_R$	109·42	109·395	163·70$_0$	163·68$_0$	215·0$_7$	215·0$_8$	267·5$_6$	267·5$_2$
differences	+0·025		+0·02$_0$		−0·0$_1$		+0·0$_4$	
$\nu'' = 3$								
mean $\delta\nu_F$	—	103·930	155·85	155·77	204·19$_5$	204·11$_4$	251·0$_{72}$	—
differences	—		+0·08		+0·08$_1$		—	
mean $\delta\nu_R$	103·965	103·930	155·81$_5$	155·7$_3$	204·2$_1$	204·23$_3$	251·0$_3$	—
differences	+0·035		+0·0$_{85}$		−0·0$_{23}$		—	

TABLE 6. SUMMARY OF $Q(m)-P(m+1)$ AND $R(m)-Q(m+1)$.
a AND *b* DIFFERENCES TAKEN SEPARATELY

0→0 BANDS

	a	*b*		*a*	*b*
$3d\,{}^3\Sigma$, $Q\,1-P\,2$	0·715	0·661	$3d\,{}^3\Pi_d$, $R\,1-Q\,2$	0·755*	0·721*
$3d\,{}^3\Pi_c$, $Q\,1-P\,2$	0·735	0·722	$3d\,{}^3\Delta_d$, $R\,1-Q\,2$	0·771	0·717
means	*0·725*	*0·692*		*0·763*	*0·719*
	+0·033			+0·044	
$3d\,{}^3\Sigma$, $R\,2-Q\,3$	0·020	0·079	$3d\,{}^3\Pi_d$, $Q\,2-P\,3$	0·033†	0·038†
$3d\,{}^3\Pi_c$, $R\,2-Q\,3$	0·037	0·019	$3d\,{}^3\Delta_d$, $Q\,2-P\,3$	0·041	0·036
means	*0·029*	*0·049*		*0·037*	*0·037*
	−0·020			0·000	
$3d\,{}^3\Pi_c$, $Q\,3-P\,4$	$0{\cdot}68_0$	$0{\cdot}62_0$			
	$+0{\cdot}06_0$				

* From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and $3d\,{}^3\Delta_d$, $Q\,2-P\,3$.
† From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and $3d\,{}^3\Delta_d$, $R\,1-Q\,2$.

1→1 BANDS

	a	*b*		*a*	*b*
$3s\,{}^3\Sigma$, $Q\,1-P\,2$	0·178	0·135	$3d\,{}^3\Pi_d$, $R\,1-Q\,2$	0·204‡	0·136
$3d\,{}^3\Sigma$, $Q\,1-P\,2$	0·191	$0{\cdot}10_1$	$3d\,{}^3\Delta_d$, $R\,1-Q\,2$	0·198	0·146
$3d\,{}^3\Pi_c$, $Q\,1-P\,2$	0·182	0·144			
means	*0·184*	*0·134*		*0·201*	*0·141*
	+0·050			+0·060	
$3s\,{}^3\Sigma$, $R\,2-Q\,3$	0·342	0·310			
$3d\,{}^3\Sigma$, $R\,2-Q\,3$	0·304	0·368	$3d\,{}^3\Pi_d$, $Q\,2-P\,3$	0·318§	0·311§
$3d\,{}^3\Pi_c$, $R\,2-Q\,3$	0·305	0·357	$3d\,{}^3\Delta_d$, $Q\,2-P\,3$	0·307	0·341
$3d\,{}^3\Delta_c$, $R\,2-Q\,3$	0·299	—			
means	*0·303*‖	*0·362*‖		*0·312*	*0·326*
	−0·059			−0·009	
$3s\,{}^3\Sigma$, $Q\,3-P\,4$	0·985	—			
$3d\,{}^3\Pi_c$, $Q\,3-P\,4$	0·949	—			
$3d\,{}^3\Delta_c$, $Q\,3-P\,4$	[0·913]	[0·833]	$3d\,\Delta_d$, $R\,3-Q\,4$	0·872	0·854
means	*0·913*	*0·833*			
	$+0{\cdot}0_8$			+0·018	

‡ From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and $3d\,{}^3\Pi\Delta_d$, $Q\,2-P\,3$.
§ From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and $R\,1-Q\,2$.
‖ Excluding $3s\,{}^3\Sigma$, $R\,2-Q\,3$.

2→2 BANDS

	a	*b*		*a*	*b*
$3d\,{}^3\Sigma$, $Q\,1-P\,2$	0·738	0·754	$3d\,{}^3\Pi_d$, $R\,1-Q\,2$	0·805¶	0·737¶
$3d\,{}^3\Pi_c$, $Q\,1-P\,2$	0·775?	0·768	$3d\,{}^3\Delta_d$, $R\,1-Q\,2$	0·783	0·743
means	*0·757?*	*0·761*		*0·794*	*0·740*
	−0·004			+0·054	
$3d\,{}^3\Sigma$, $R\,2-Q\,3$	0·896	0·902	$3d\,{}^3\Pi_d$, $Q\,2-P\,3$	0·926**	0·891**
$3d\,{}^3\Pi_c$, $R\,2-Q\,3$	0·901	d.i.	$3d\,{}^3\Delta_d$, $Q\,2-P\,3$	0·882	0·903
$3d\,{}^3\Delta_c$, $R\,2-Q\,3$	0·890	$0{\cdot}91_0$			
means	*0·896*	$0{\cdot}90_6$		$0{\cdot}90_4$	*0·897*
	$-0{\cdot}01_0$			$+0{\cdot}00_7$	
$3d\,{}^3\Pi_c$, $Q\,3-P\,4$	*0·301*	*0·249*			
	+0·052				

¶ From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and mean $\Pi_d\Delta_d$, $Q\,2-P\,3$.
** From $3d\,{}^3\Pi_d$, $R\,1-P\,3$ and $3d\,{}^3\Delta_d$, $R\,1-Q\,2$.

TABLE 6 (*continued*)

3→3 BANDS

	a	b		a	b
$3d\,^3\Sigma$, $Q\,1-P\,2$	0·520	d.i.	$3d\,^3\Pi_d$, $R\,1-Q\,2$	0·504††	0·527††
			$3d\,^3\Delta_d$, $R\,1-Q\,2$	0·512	0·496
means	*0·520*	—		*0·508*	*0·511*
				−0·003	
$3d\,^3\Sigma$, $R\,2-Q\,3$	0·679	0·685	$3d\,^3\Pi_d$, $Q\,2-P\,3$	0·659‡‡	0·664‡‡
			$3d\,^3\Delta_d$, $Q\,2-P\,3$	0·663	0·648
means	*0·679*	*0·685*		*0·661*	*0·656*
	−0·006			+0·005	

†† From $3d\,^3\Pi_d$, $R\,1-P\,3$ and $3d\,^3\Delta_d$, $Q\,2-P\,3$.
‡‡ From $3d\,^3\Pi_d$, $R\,1-P\,3$, and mean $\Pi_d\Delta_d$, $R\,1-Q\,2$.

TABLE 7. SUMMARY OF $Q'(m)-P'(m+1)$ AND $R'(m)-Q'(m+1)$
a AND *b* DIFFERENCES TAKEN SEPARATELY

0→0 BANDS

	a	b		a	b
$3d\,^3\Sigma$, $R'1-Q'2$	0·748	0·791			
$3d\,^3\Pi_c$, $R'1-Q'2$	?0·785	—			
$3d\,^3\Delta_c$, $R'1-Q'2$	0·742	0·773			
means	*0·745*	*0·782*			
	−0·037				
$3d\,^3\Delta_c$, $Q'2-P'3$	*0·968*	—			

1→1 BANDS

	a	b
$3d\,^3\Sigma$, $R'1-Q'2$	0·980	—
$3d\,^3\Pi_c$, $R'1-Q'2$	0·972*	0·046†
$3d\,^3\Delta_c$, $R'1-Q'2$	0·980	—
means	*0·977*	*0·046*
	−0·069	

* From $3d\,^3\Pi_c$, $R'1-P'3 = 0{\cdot}698$ and $3d\,^3\Delta_c$, $Q'2-P'3 = 0{\cdot}726$.
† From $3d\,^3\Pi_c$, $R'1-P'3 = 0{\cdot}714$ and $3d\,^3\Delta_c$, $Q'2-P'3 = 0{\cdot}668$.

2→2 BANDS

	a	b		a	b
$3d\,^3\Sigma$, $R'1-Q'2$	0·398	[0·416]	$3d\,^3\Pi_d$, $Q'1-P'2$	*$0{\cdot}42_1$*	*$[0{\cdot}38_1]$*
$3d\,^3\Delta_c$, $R'1-Q'2$	0·371	0·450			
means	*0·384*	*[0·433]*			
	−[0·049]				
				$+[0{\cdot}04_0]$	
			$3d\,^3\Delta_d$, $R'2-Q'3$	0·716	0·686
$3d\,^3\Delta_c$, $R'3-Q'4$	*0·096?*	*0·002?*		+0·030	
	+0·094?				

3→3 BANDS

	a	b
$3d\,^3\Pi_c$, $R'1-Q'2$	0·872‡	0·893§
$3d\,^3\Delta_c$, $R'1-Q'2$	0·908	0·911
means	*0·890*	*0·902*
	−0·012	
$3d\,^3\Delta_c$, $Q'2-P'3$	*0·778*	*0·748*
	+0·030	

‡ From $3d\,^3\Pi_c$, $R'1-P'3 = 0{\cdot}650$ and $3d\,^3\Delta_c$, $Q'2-P'3 = 0{\cdot}778$.
§ From $3d\,^3\Pi_c$, $R'1-P'3 = 0{\cdot}64_1$ and $3d\,^3\Delta_c$, $Q'2-P'3 = 0{\cdot}748$.

In each band the value is highest for $\overline{R1-Q2}\,a-\overline{Q1-P2}\,a$ and falls steadily as we proceed through

$$\overline{R1-Q2}\,b-\overline{Q1-P2}\,b \quad \text{and} \quad \overline{Q2-P3}\,a-\overline{R2-Q3}\,a \quad \text{to} \quad \overline{Q2-P3}\,b-\overline{R2-Q3}\,b.$$

The average of all four values is highest for the 0 → 0 bands and lowest for the 3 → 3 bands, where it becomes negative. In the 0 → 0 bands three values are positive and one negative, in the 1 → 1 and 2 → 2 bands two are positive and two negative, and in the 3 → 3 bands three are negative and one unknown.

In the dashed transitions it is likely that something like this occurs with the sign reversed, but the data are too scanty to test it satisfactorily.

(7) The data are very restricted when one wishes to use them to combine the results for the dashed transitions with those for the undashed. There is, however, one dashed rotational difference which follows right through the vibrational sequence 0 → 0 to 3 → 3, namely, that of the $R'1-Q'2$ values. If these are compared with the next mean $\overline{R2-Q3}$ values for the same $v'v''$ the results shown in table 8 are obtained. This shows that, at least so far as this one available example is concerned, when the data are combined in this way: (*a*) the alternations with successive vibrational quantum numbers disappear, (*b*) the gap between the *a* and *b* differences is much reduced, and (*c*) it tends to disappear with increase in the vibrational quantum number.

TABLE 8

	0 → 0		1 → 1		2 → 2		3 → 3	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
$R'1-Q'2$	·745	·782	·977	·046	·384	[·433]	·890	·902
$R2-Q3$	·029	·049	·303	·362	·896	$·90_6$	·679	·685
difference	0·716	0·733	0·674	0·684	0·488	$[0·52_7]$	0·211	0·217
	0·017		0·010		$[0·03_9]$		0·006	

The discrepant value from the 2 → 2 band may arise from the *b* value for which the calculated $R'1-Q'2$ difference and the observed $R2-Q3$ are probably less exact than the others.

(8) The value ·079 of the decimal part of the $3d\,^3\Sigma$, 0 → 0, $R2-Q3\,b$ difference (table 6) seems high. It is just possible that the *b* component of the $R2$ line is split by a perturbation and that the unassigned structure (*n*) (wave number ·052, intensity ratio 1: see table 3 of Paper I) is really part of it. In that case the value 0·079 would be reduced, possibly to as little as 0·043, depending on the exact value of the intensity ratios of the components (*l*), (*m*) and (*n*) of 16834·27 (10) *G*. If this interpretation of the pattern could be established, important results would follow.

Having completed what may be termed the empirical analysis of the data, we shall now proceed to attack the problem in another direction.

3. THE DOUBLET SEPARATIONS

It is evident from tables 1 and 2 that the majority of the lines starting from $3s\,{}^3\Sigma$ and the complex $3d\,{}^3\Sigma, \Pi_{cd}, \Delta_{cd}$ and ending on $2p\,{}^3\Pi_{cd}$ are each split with the resolution of the echelon into two components a and b, a having the higher frequency and also the greater intensity. The evidence is complete enough to be convincing that this statement applies to all lines ending on $2p\,{}^3\Pi_{cd}$. Where only one component appears, this is always due either to the conditions not being sensitive enough to make the weaker component visible, or, owing to the doublet separation being different for different lines, in a relatively small number of cases the components are not clearly separated with our resolution, the doublets appearing as single lines of abnormal width.

The largest doublet separation we have observed is about 0·22 cm.$^{-1}$, the smallest about one-third of this amount.

From an observational point of view, the lines (and doublets) divide themselves into two groups according to the properties of the components. We have called these the regular and irregular group respectively. There is no very sharp rule of demarcation for every line, but most of them can be picked out by inspection. The regular group has a considerably greater average doublet separation. There are several well-authenticated examples with $\Delta\nu$ not much under 0·22 cm.$^{-1}$, very few have $\Delta\nu$ less than about 0·14, and it is doubtful if the measures for any of these are to be relied on. In the irregular group there are no reliable $\Delta\nu$'s greater than 0·18, but values as low as 0·06 and 0·08 seem well secured. The two groups also differentiate themselves by the values of the intensities of the components. In the regular group, the estimated values of $I(a/b)$ practically all lie within the limits 3·0 and 1·9, most of them close to 2·4. The four values outside the limits 3·0 and 1·9 are probably all doubtful. The range of $I(a/b)$ in the irregular group is much greater. There are several well-secured values at or close to 5·0 and 0·8 and a few cases, not very secure, of values outside these limits.

It will be shown below that the division into the regular and irregular group can be related to the symmetry properties of the states $2p\,{}^3\Pi_{cd}$.

Before going further, it will be convenient to have before us a table of all the splittings and intensity ratios which have been observed. These, followed by our estimated weights, for $\Delta\nu$ and $I(a/b)$, together with the classified transitions and ν_G values are assembled in table 9.

The following discussion is mainly a first-order treatment which neglects any fine structure that may exist in the upper states $3s\,{}^3\Sigma$ and $3d\,{}^3\Sigma, \Pi_{cd}, \Delta_{cd}$ and attributes all the effects so far observed to the fine-structure components of the lower states $2p\,{}^3\Pi_{cd}$. This mode of treatment is justified as a first approximation by the following considerations:

(1) Almost all the lines which so far have been found to have any directly perceptible fine structure are lines which end on $2p\,{}^3\Pi_{cd}$. The number of others is no greater than might be expected to arise from unidentified blends.

TABLE 9. SPLITTINGS AND INTENSITY RATIOS FOR $3s\,^3\Sigma$ AND $3d\sigma\,^3\Sigma\;^3\Pi_{cd}\,^3\Delta_{cd}$ COMPLEX $\rightarrow 2p\,^3\Pi_{cd}$

$0' \rightarrow 0''$ REGULAR

line	$d^3\Sigma, Q1$	$\Pi_d, R1$	$\Pi_c, Q1$	$\Delta_d, R1$
$\Delta\nu$ (weight)	0·194 (6)	0·216 (4)	?0·185 (1)	0·213 (3)
$I(a/b)$ (weight)	2·5 (3)	2·5 (2)	?2·0 (1)	2·5 (2)
ν_G	16854·86	17199·28	17212·01	17572·31

Line	$d^3\Sigma, Q'2$	$\Pi_c, Q'2$	$\Delta_c, Q'2$
$\Delta\nu$ (weight)	0·189 (5)	0·18 (0)	0·160 (3)
$I(a/b)$ (weight)	2·95 (6)	2·5 (0)	2·5 (1)
ν_G	16764·12	17248·55	17467·13

line	$d^3\Sigma, Q3$	$\Pi_d, P3$	$\Pi_d, R3$	$\Pi_c, Q3$	$\Delta_d, P3$	$\Delta_c, Q3$
$\Delta\nu$ (weight)	0·173 (6)	0·177 (3)	0·160 (4)	0·151 (3)	0·144 (3)	0·216 (2)
$I(a/b)$ (weight)	2·5 (2)	2·5 (2)	2·5 (1)	2·5 (2)	2·5 (1)	2·5 (1)
ν_G	16654·24	16898·48	17199·28	17261·34	17271·56	17537·02

no $K'' = 4$ levels

line	$\Pi_c, Q5$
$\Delta\nu$ (weight)	?0·15 (0)
$I(a/b)$ (weight)	?2 (1)
ν_G	17252·65

$0' \rightarrow 0''$ IRREGULAR

line	$d^3\Sigma, R'1$	$\Pi_d, Q'1$	$\Pi_c, R'1$	$\Delta_c, R'1$
$\Delta\nu$ (weight)	0·146 (1)	??0·20 (1)	—	0·129 (2)
$I(a/b)$ (weight)	4·5 (0)	?5 (0)	⩾5 (0)	4·0 (1)
ν_G	16884·94	17125·07	17369·32	17588·00

line	$d^3\Sigma, P2$	$d^3\Sigma, R2$	$\Pi_c, P2$	$\Pi_c, R2$	$\Delta_d, Q2$
$\Delta\nu$ (weight)	0·140 (1)	0·114 (3)	0·172 (3)	0·169 (2)	0·149 (3)
$I(a/b)$ (weight)	?2·0 (0)	4·0 (3)	4·0 (3)	4·0 (0)*	?4·5 (1)
ν_G	16734·20	16834·27	17091·29	17441·32	17451·59

line	$d^3\Sigma, R'3$	$\Pi_c, P'3$	$\Delta_c, P'3$
$\Delta\nu$ (weight)	0·12 (1)	?0·10 (1)	?0·16 (0)
$I(a/b)$ (weight)	4·0 (0)	?1·0 (0)	—
ν_G	16774·97	17068·60	17287·27

line	$d^3\Sigma, R4$	$\Pi_c, P4$	$\Delta_d, Q4$	$\Delta_c, P4$
$\Delta\nu$ (weight)	0·06 (3)	0·09 (1)	?0·08 (1)	0·08 (3)
$I(a/b)$ (weight)	1·0 (3)	?1·0 (1)	?1·5 (1)	?0·60 (2)
ν_G	16710·63	17023·67	17527·44	17299·30

* This might be $\sim 2·5:1$ if q (Part I, table 3) is of negligible intensity.

$1' \rightarrow 1''$ REGULAR

line	$s^3\Sigma, Q1$	$d^3\Sigma, Q1$	$\Pi_d, R1$	$\Pi_c, Q1$	$\Delta_d, R1$
$\Delta\nu$ (weight)	0·20 (1)	0·196 (5)	0·205 (6)	0·220 (2)	0·190 (6)
$I(a/b)$ (weight)	?2·5 (1)	2·5 (2)	2·5 (2)	?2·0 (0)	1·9 (5)
ν_G	16917·50	16603·20	16978·52	16988·72	17426·25

line	$s^3\Sigma, Q'2$	$d^3\Sigma, Q'2$	$\Pi_d, R'2$	$\Delta_c, Q'2$
$\Delta\nu$ (weight)	?0·162 (1)	0·179 (5)	0·229 (0)	0·161 (4)
$I(a/b)$ (weight)	?3·0 (0)	2·5 (1)	2·5 (0)	2·5 (2)
ν_G	16916·57	16515·45	16988·72	17316·56

line	$s^3\Sigma, Q3$	$d^3\Sigma, Q3$	$\Pi_d, R3$	$\Pi_d, P3$	$\Pi_c, Q3$	$\Delta_d, P3$	$\Delta_d, R3$	$\Delta_c, Q3$
$\Delta\nu$ (weight)	0·18 (1)	0·182 (2)	?0·21 (0)	0·146 (4)	0·170 (4)	0·172 (2)	?0·152(1)	0·200 (2)
$I(a/b)$ (weight)	?2·5 (1)	2·5 (2)	3·0 (0)	2·5 (1)	2·5 (2)	?2·5 (0)	?3·0 (0)	2·5 (1)
	16915·33	16414·37	16990·71	16692·00	17047·15	17139·77	17591·22	17357·64

no $K'' = 4$ lines

line	$\Delta_d, P5$
$\Delta\nu$ (weight)	?0·08 (1)
$I(a/b)$ (weight)	?1·0 (0)
ν_G	17084·56

TABLE 9 (*continued*)

$1' \rightarrow 1''$ IRREGULAR

line	$d^3\Sigma, R'1$	$\Pi_d, Q'1$	$\Pi_c, R'1$
$\Delta\nu$ (weight)	—	—	0·15 (0)
$I(a/b)$ (weight)	$\geqslant$5·0 (2)	—	5·0 (0)
ν_G	16630·47	16899·76	17138·90

line	$s^3\Sigma, R2$	$s^3\Sigma, P2$	$d^3\Sigma, R2$	$d^3\Sigma, P2$	$\Pi_c, P2$	$\Pi_c, R2$	$\Delta_d, Q2$	$\Delta_c, R2$
$\Delta\nu$ (weight)	?0·225 (0)	0·150 (3)	0·118 (3)	0·106 (2)	0·182 (3)	0·118 (2)	0·138 (2)	—
$I(a/b)$ (weight)	?3·0 (0)	5·0 (4)	5·0 (4)	3·0 (2)	5·0 (3)	2 (0)	?6·0 (1)	$\geqslant$4 (0)
ν_G	17086·67	16802·37	16584·72	16488·15	16873·46	17218·50	17311·08	17529·00

line	$d^3\Sigma, R'3$	$\Pi_c, P'3$	$\Delta_c, P'3$
$\Delta\nu$ (weight)	0·12 (1)	0·166 (3)	0·103 (2)
$I(a/b)$ (weight)	3·5 (1)	3·5 (2)	4·0 (2)
ν_G	16530·23	16852·22	17144·93

line	$d^3\Sigma, R4$	$\Pi_c, P4$	$\Delta_d, Q4$
$\Delta\nu$ (weight)	0·100 (3)	?0·12 (1)	0·134 (3)
$I(a/b)$ (weight)	0·8 (5)	?1·0 (0)	?1·0 (0)
ν_G	16469·93	16820·29	17364·30

$2' \rightarrow 2''$ REGULAR

line	$d^3\Sigma, Q1$	$\Pi_d, R1$	$\Pi_c, Q1$	$\Delta_d, R1$
$\Delta\nu$ (weight)	0·157 (1)	0·189 (3)	?0·124 (1)	0·185 (5)
$I(a/b)$ (weight)	?3·0 (0)	?2·5 (0)	?3·0 (0)	2·5 (2)
ν_G	16297·25	16730·55	16684·88	17279·01

line	$d^3\Sigma, Q'2$	$\Pi_d, P'2$	$\Delta_c, Q'2$	$\Delta_d, R'2$
$\Delta\nu$ (weight)	0·168 (3)	0·11 (0)	0·175 (3)	0·138 (2)
$I(a/b)$ (weight)	2·5 (2)	3·0 (1)	3·0 (2)	?2·5 (1)
ν_G	16222·72	16538·98	17171·27	17348·12

line	$d^3\Sigma, Q3$	$\Pi_d, P3$	$\Pi_d, R3$	$\Pi_c, Q3$	$\Delta_d, P3$	$\Delta_c, Q3$
$\Delta\nu$ (weight)	0·166 (3)	0·154 (5)	0·154 (3)	0·170 (1)	0·166 (3)	0·150 (4)
$I(a/b)$ (weight)	2·5 (2)	2·5 (2)	1·5 (2)	?2·5 (0)	?2·5 (1)	2·5 (2)
ν_G	16132·09	16458·03	16753·44	16734·20	17006·43	17192·31

line	$\Delta_c, Q'4$							$\Delta_c, P5$
$\Delta\nu$ (weight)	0·118 (1)							0·140 (1)
$I(a/b)$ (weight)	2·0 (0)							?2·0 (1)
ν_G	17218·50							16934·41

$2' \rightarrow 2''$ IRREGULAR

line	$d^3\Sigma, R'1$	$\Pi_d, Q'1$	$\Pi_c, R'1$	$\Delta_c, R'1$
$\Delta\nu$ (weight)	0·15 (0)	0·15 (0)	?0·17 (1)	0·096 (2)
$I(a/b)$ (weight)	—	$\nless$4 (0)	?2·5 (0)	?6·0 (5)
ν_G	16332·11	16648·40	16822·00	17280·68

line	$d^3\Sigma, P2$	$d^3\Sigma, R2$	$\Pi_c, P2$	$\Pi_c, R2$	$\Delta_d, Q2$	$\Delta_c, R2$
$\Delta\nu$ (weight)	0·173 (1)	0·160 (1)	0·167 (1)	0·15 (0)	0·145 (2)	0·130 (1)
$I(a/b)$ (weight)	?4·0 (0)	?4·0 (0)	4·0 (2)	$\geqslant$4·0 (0)	5·0 (1)	4·0 (1)
ν_G	16187·50	16295·02	16575·12	16897·19	17169·22	17355·14

line	$\Delta_d, Q'3$	$\Delta_c, R'3$
$\Delta\nu$ (weight)	?0·108 (1)	?0·212 (0)
$I(a/b)$ (weight)	?3·0 (0)	?3·0 (0)
ν_G	17184·46	17433·50

line	$d^3\Sigma, R4$	$\Pi_c, P4$	$\Delta_c, P4$
$\Delta\nu$ (weight)	0·121 (2)	0·118 (3)	—
$I(a/b)$ (weight)	1·5 (1)	0·9 (2)	$\geqslant$0·6 (0)
ν_G	16192·25	16517·89	16976·03

TABLE 9 (*continued*)

$3' \rightarrow 3''$ REGULAR

line	$d^3\Sigma$, $Q1$	Π_d, $R1$	Π_c, $Q1$	Δ_d, $R1$
$\Delta\nu$ (weight)	0·166 (3)	0·124 (1)	0·211 (2)	0·182 (6)
$I(a/b)$ (weight)	2·5 (1)	2·5 (1)	2·5 (0)	2·5 (2)
ν_G	15931·94	16445·80	16388·32	17129·89

line	Π_d, $R'2$	Δ_d, $R'2$	Δ_c, $Q'2$
$\Delta\nu$ (weight)	0·174 (1)	??0·136 (1)	0·179 (1)
$I(a/b)$ (weight)	?1·0 (1)	??2·5 (0)	?2·5 (0)
ν_G	16465·55	17186·00	17026·36

line	$d^3\Sigma$, $Q3$	Π_d, $P3$	Π_d, $R3$	Δ_d, $P3$	Δ_c, $Q3$
$\Delta\nu$ (weight)	0·141 (2)	0·132 (2)	0·169 (1)	0·151 (3)	0·160 (2)
$I(a/b)$ (weight)	?2·0 (1)	?3·0 (0)	2·5 (0)	2·0 (1)	2·5 (0)
ν_G	15795·05	16186·63	16476·10	16870·70	17033·21

line	Δ_d, $P'4$	Δ_c, $Q'4$	$d^3\Sigma$, $Q5$
$\Delta\nu$ (weight)	?0·110 (0)	?0·128 (2)	0·114 (1)
$I(a/b)$ (weight)	$\geqslant$4·0 (0)	?2·0 (0)	?3·0 (1)
ν_G	16825·98	17040·75	15610·91

$3' \rightarrow 3''$ IRREGULAR

line	Π_c, $R'1$	Δ_c, $R'1$
$\Delta\nu$ (weight)	0·150 (1)	0·176 (1)
$I(a/b)$ (weight)	?3·0 (0)	5·0 (>0)
ν_G	16511·46	[17130·30]

ine	$d^3\Sigma$, $R2$	$d^3\Sigma$, $P2$	Δ_d, $Q2$
$\Delta\nu$ (weight)	0·135 (2)	0·15 (0)	0·166 (1)
$I(a/b)$ (weight)	1·5 (2)	4·0 (0)	?5·0 (0)
ν_G	15949·72	15827·48	17025·41

line	Π_c, $P'3$	Δ_c, $P'3$
$\Delta\nu$ (weight)	0·141 (1)	—
$I(a/b)$ (weight)	4·0 (0)	—
ν_G	16251·79	16870·70

line	Δ_d, $Q4$
$\Delta\nu$ (weight)	0·167 (1)
$I(a/b)$ (weight)	1·0 (0)
ν_G	[17033·66]

(2) The doublet separations are roughly the same for all lines going down to all levels of $2p\,{}^3\Pi_c$ having the same ν'' and K''. The relative intensities of the two components are also characteristic of the same variables to the same kind of approximation.

(3) The same statements can be made regarding all lines going down to the same ν'' and K'' of $2p\,{}^3\Pi_d$. The values for the $2p\,{}^3\Pi_d$ levels are, however, not the same as for the $2p\,{}^3\Pi_c$ levels, but the values are characteristic for the respective levels.

(4) We have some evidence, to be described in subsequent papers, of the existence of fine structure in states of H_2 other than $2p\,{}^3\Pi_{cd}$. In none of the cases so far examined is it, however, on a scale exceeding one-tenth of that of the fine structure in $2p\,{}^3\Pi_{cd}$.

In this connexion it is as well to remember that evidence of the absence of fine structure is often more convincing than evidence of its existence.

In view of (4) we shall, however, pay some regard to the possibility of fine structure in the upper levels where this can be done fairly simply and in a reasonable compass.

4. THE TWO GROUPS

While the division into a 'regular' and an 'irregular' group has an observational basis, this does not, as we have pointed out, sharply define the contents of the groups. To do this, it is necessary to consider the transitions from $3s\,^3\Sigma$ and the complex $3d\,^3\Sigma$, $3d\,^3\Pi_c$, $3d\,^3\Pi_d$, $3d\,^3\Delta_c$, $3d\,^3\Delta_d$ to $2p\,^3\Pi_c$ and $2p\,^3\Pi_d$. A diagram of these is shown in figure 1 (so far as this discussion is concerned the state labelled $d\,^3\Sigma$ will do equally well for both $3d\,^3\Sigma$ and $3s\,^3\Sigma$). Both $3d\,^3\Sigma$ and $3s\,^3\Sigma$ are even (Σg) states; $2p\,^3\Pi_c$, $3d\,^3\Pi_c$ and $3d\,^3\Delta_c$ also have this symmetry. $2p\,^3\Pi_d$, $3d\,^3\Pi_d$ and $3d\,^3\Delta_d$ are odd (u) states. The transitions which go down to $2p\,^3\Pi_c$ are limited to those which satisfy the selection rules $g \rightarrow g$ and $\Delta K = 0$ or $u \rightarrow g$ and $\Delta K = \pm 1$, those to $2p\,^3\Pi_d$ satisfy $u \rightarrow u$ and $\Delta k = 0$ or $g \rightarrow u$ and $\Delta K = \pm 1$. Omitting the prefix $3d^3$, the lines which go down to the three lowest rotational levels of $2p\,^3\Pi_c$ are $\Sigma Q1$, $\Pi_c Q1$, $\Pi_d R1$, $\Delta_d R1$, $\Sigma Q'2$, $\Pi_c Q'2$, $\Pi_d P'2$, $R'2$, $\Delta_c Q'2$, $\Delta_d R'2$, $\Sigma Q3$, $\Pi_c Q3$, $\Pi_d P3$, $R3$, $\Delta_c Q3$ and $\Delta_d P3$, $R3$, whilst those going down to $2p\,^3\Pi_d$ are $\Sigma P'1$, $R'1$, $\Pi_d Q'1$, $\Pi_c R'1$, $\Delta_c R'1$, $\Sigma P2$, $R2$, $\Pi_c P2$, $R2$, $\Pi_d Q2$, $\Delta_c R2$, $\Delta_d Q2$, $\Sigma P'3$, $R'3$, $\Pi_c P'3$, $R'3$, $\Pi_d Q'3$, $\Delta_c P'3$, $R'3$ and $\Delta_d Q'3$. It will be seen from table 9 that the transitions down to $2p\,^3\Pi_c$ are just those which form the regular group, and the transitions down to $2p\,^3\Pi_d$ the irregular group.

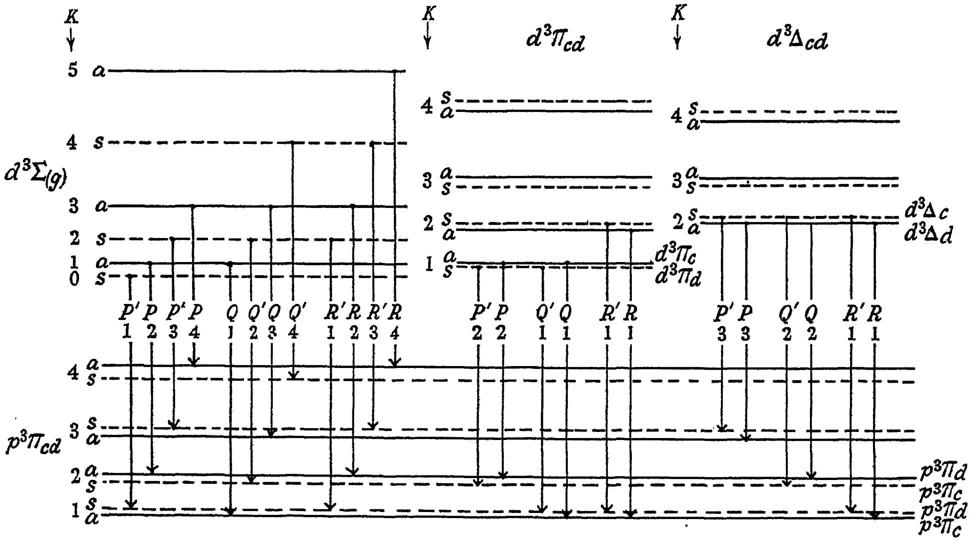


FIGURE 1. The rotational structure of $d^3\Sigma_{(g)}$, $d^3\Pi_{cd}$, $d^3\Delta_{cd}$ and $p^3\Pi_{cd}$ levels and the transitions between them.

To the extent that the separation of the upper levels can be considered negligible all the measured separations will be due to the lower final levels. Within this limitation the properties of the regular group are those of $2p\,{}^3\Pi_c$ and of the irregular group they are those of $2p\,{}^3\Pi_d$.

It will be seen from table 9 that the separations $\Delta\nu$ are fairly constant for the regular doublets. In the $0 \rightarrow 0$, $1 \rightarrow 1$, and $2 \rightarrow 2$ bands, however, there is evidence of a tendency of $\Delta\nu$ to fall slowly as K'' increases from 0 to 3 and this is followed by a sharp drop at K'' about 4 or 5. In the $3 \rightarrow 3$ bands this tendency of $\Delta\nu$ to change with K'' seems less marked. For the irregular doublets the $\Delta\nu$'s suggest similar tendencies but the average separations are much smaller. The reduction of $\Delta\nu$ is mainly at the low K'' end where the values are highest.

In the regular doublets the intensity ratio $I(a/b)$ shows little systematic variation from about 2·5, but here again there seems to be a suggestion of a slow fall from $K'' = 1$ to 3 followed by a sharper drop at $K'' = 4$ for $\nu'' = 0$ to 2 with the $3 \rightarrow 3$ band behaving differently. In the irregular doublets $I(a/b)$ falls from about 5·0 to about 1·0 as K'' increases from 1 to 4 at each value of ν''.

There is some difficulty in assessing the true trend of the real values by inspection of a compilation like table 9, as there is a very great variation in the reliability of the different determinations. This is shown by the fact that the weights, both of the $\Delta\nu$'s and of the $I(a/b)$, cover the whole possible range (0) to (6) assignable.

We have endeavoured to eliminate this by forming a weighted mean of each determination of the same quantity by weighting them in the proportion of the weight assigned in table 9 and giving the result a weight equal to the sum of the original weights. The final weights will give some sort of basis for comparing the weight of evidence behind all the data. The result of this is given in table 10. We will leave the numbers in this table to speak for themselves.

5. INFLUENCE OF THE UPPER STATES ON THE DOUBLET SEPARATION

If there were no fine structure in the lines, the energies of the initial and final states could be expressed respectively by $F'(v', k')$ and $F''(\nu'', K'')$ where ν', K' and ν'', K'' are the vibrational rotational quantum numbers of the levels. As we have seen the final states at least are double so that each $F''(\nu'', K'')$ has two values which we may denote by $F_e(\nu''K'')$ and $F''_f(\nu''K'')$, $F''_e < F''_f$. This will double the number of lines, each single line now being replaced by a doublet with separation $F''_f - F''_e$.

Confining discussion to the simplest case likely to be detectable with the resolution at present known, suppose the upper levels are also double with energies expressed by $F'_e(\nu'K')$ and $F'_f(\nu'K')$, $F'_f > F'_e$. Each of the previous doublets will now become another doublet, the whole forming a quadruplet whose components have frequencies, omitting the ν''s and K''s for brevity, (1) $F'_f - F''_e$, (2) $F'_e - F''_e$, (3) $F'_f - F''_f$, (4) $F'_e - F''_f$ in the order of ν decreasing. With the present resolution, this quadruplet will appear as a doublet, and the mean frequencies of the two components will depend on the transition probabilities between the upper and lower fine-structure levels e and f.

TABLE 10. MEAN SPLITTINGS AND INTENSITY RATIOS FOR TRANSITIONS TO $2p\,^3\Pi$

REGULAR

v'' / K''	0	1	2	3	0	1	2	3	mean intensity ratio
	splittings				intensity ratios				
1	*0·203 (18)*	*0·200 (25)*	*0·181 (14)*	*0·176 (16)*	*2·4 (12)*	*2·2 (15)*	*2·3 (6)*	*2·5 (8)*	*2·3*$_3$
	0·204 (14)	0·199 (20)	0·182 (10)	0·178 (12)	2·4$_4$ (8)	2·2 (10)	2·5 (2)	2·5 (4)	2·3$_6$
	0·204 (10)	*0·199 (15)*	*0·186 (6)*	*0·181 (8)*	*2·5 (4)*	*2·1 (6)*	*2·5 (1)*	*2·5 (1)*	*2·3*$_0$
2	*0·178 (11)*	*0·174 (14)*	*0·158 (12)*	*0·163 (6)*	*2·8*$_2$ *(10)*	*2·5*$_7$ *(7)*	*2·7*$_5$ *(10)*	*1·75 (4)*	*2·60*
	0·178 (8)	0·170 (10)	0·163 (8)	0·163 (3)	2·8$_8$ (7)	2·5 (3)	2·7$_5$ (6)	1·0 (1)	2·6$_6$
	0·179 (6)	*0·171 (7)*	*0·165 (5)*	—	*2·9*$_5$ *(5)*	*2·5 (1)*	*2·5 (1)*	—	*2·8*$_2$
3	*0·168 (27)*	*0·171 (24)*	*0·158 (25)*	*0·149 (15)*	*2·5 (15)*	*2·5*$_7$ *(15)*	*2·3 (15)*	*2·0 (7)*	*2·4*$_0$
	0·168 (21)	0·169 (16)	0·157 (19)	0·149 (10)	2·5 (9)	2·5 (7)	2·2$_8$ (9)	2·0 (2)	2·3$_9$
	0·167 (15)	*0·167 (9)*	*0·157 (13)*	*0·147 (5)*	*2·5 (3)*	*2·5 (2)*	*2·2*$_5$ *(4)*	—	*2·4*$_0$
4	—	—	*0·118 (2)*	*0·121 (5)*	—	—	*2·0 (1)*	*3·0 (2)*	*2·6*$_6$
	—	—	0·118 (1)	0·122 (3)	—	—	—	—	—
	—	—	—	*0·128 (1)*	—	—	—	—	—
5	*0·15 (1)*	*0·08 (2)*	*0·140 (2)*	*0·114 (2)*	*2·0 (2)*	*1·0 (1)*	*2·0 (2)*	*3·0 (2)*	*2·1*$_4$
	—	0·08 (1)	0·140 (1)	0·114 (1)	2·$_0$ (1)	—	2·$_0$ (1)	3·0 (1)	2·3$_3$
	—	—	—	—	—	—	—	—	—
					IRREGULAR				
1	*0·15*$_2$ *(7)*	*0·15 (1)*	*0·134 (7)*	*0·163 (4)*	*4·5 (5)*	*⩾5·0 (4)*	*5·3 (8)*	*4·0 (2)*	*4·9*
	0·15 (4)	—	0·121 (3)	0·163 (2)	4·2$_5$ (2)	⩾5·$_0$ (2)	6·0 (5)	—	5·4
	0·129 (1)	—	0·096 (1)	—	—	⩾5·$_0$ (1)	6·0 (4)	—	*5·8*
2	*0·149 (17)*	*0·141 (22)*	*0·154 (12)*	*0·148 (6)*	*3·9*$_1$ *(12)*	*4·6 (21)*	*4·2 (9)*	*2·2 (5)*	*4·1*
	0·149 (12)	0·138 (15)	0·153 (6)	0·145 (3)	4·0$_7$ (7)	4·8 (14)	4·2$_5$ (4)	1·5 (2)	4·2$_8$
	0·148 (7)	*0·140 (9)*	*0·145 (1)*	*0·135 (1)*	*4·0 (4)*	*4·8 (9)*	*4·0 (1)*	*1·5 (1)*	*4·3*
3	*0·12 (5)*	*0·135 (9)*	*0·143 (3)*	*0·141 (2)*	*2·5 (2)*	*3·7 (8)*	*3·0 (2)*	*4·0 (1)*	*3·4*
	0·11 (2)	0·137 (6)	0·108 (1)	0·141 (1)	—	3·7 (5)	—	—	3·7
	—	*0·145 (3)*	—	—	—	*3·7*$_5$ *(2)*	—	—	*3·7*$_5$
4	*0·07*$_5$ *(12)*	*0·117 (10)*	*0·119 (7)*	*0·167 (2)*	*1·0*$_8$ *(11)*	*0·8*$_5$ *(8)*	*0·9*$_6$ *(5)*	*1·0 (1)*	*0·9*$_8$
	0·07$_4$ (8)	0·117 (7)	0·119 (5)	0·167 (1)	0·9$_6$ (7)	0·8 (5)	1·1 (3)	—	0·9$_4$
	0·07$_0$ *(4)*	*0·117 (4)*	*0·119 (3)*	—	*0·8*$_6$ *(3)*	*0·8 (4)*	*0·9 (1)*	—	*0·8*$_4$

For each K'' and ν'' three values of splitting and three values of intensity ratio are given. The middle values are obtained when the normal weightings for these quantities, given in table 9, are used. The top values are those obtained when the normal weights are increased by 1 and the bottom values those obtained when the normal weights are decreased by 1. Thus, the top values are weighted means of all observations good and bad whilst the bottom values include the best observations only—i.e. those with weight ⩾ 2. These changes in the method of averaging have little effect on the final result.

TABLE 11

	regular levels				irregular levels		
v' v''	$3d\,^3\Pi_d$, $R1$, $P3$	$3d\,^3\Delta_d$, $R1$, $P3$	$3d\,^3\Delta_d$, $R'2$, $P'4$	$3d\,^3\Delta_d$, $R3$, $P5$	$3d\,^3\Pi_c$, R', $P'3$	$3d\,^3\Delta_c$, $R'1$, $P'3$	$3d\,^3\Pi_c$, $R2$, $P4$
0′ 0″	+0·039	—	—	—	—	?−0·031	+0·079
1′ 1″	+0·059	+0·018	—	?+0·072	−0·016	—	?−0·002
2′ 2″	+0·035	+0·021	—	—	—	—	+0·032
3′ 3″	?−0·008	+0·031	?+0·026	—	+0·009	—	—

If the transitions are $e \rightarrow f$, $e \rightarrow e$ and $f \rightarrow e$, $f \rightarrow f$, the transitions of each pair having equal probability the observed doublet will have components of frequencies

$$\tfrac{1}{2}(F'_f - F''_e + F'_e - F''_e) \quad \text{and} \quad \tfrac{1}{2}(F'_f - F''_f + F'_e - F''_f)$$

and the doublet width will still be $F''_f - F''_e$ (*a*). If, however, the transitions are $f \rightleftarrows e$ and $\begin{matrix} e \nrightarrow e \\ f \nrightarrow f \end{matrix}$, the component frequencies are $F'_f - F''_e$ and $F'_e - F''_f$ and the doublet width $F''_f - F''_e + F'_f - F'_e$ (*b*); if they are $\begin{matrix} f \rightarrow f \\ e \rightarrow e \end{matrix}$ and $\begin{matrix} f \nrightarrow e \\ e \nrightarrow f \end{matrix}$, the component frequencies are $F'_f - F''_f$ and $F'_e - F''_e$ and the doublet width $F'_f - F'_e - (F''_f - F''_e)$ (*c*). In general the doublet width observed can have any value between (*b*) the sum of the separations due to the energy differences of the initial and final levels and (*c*) the differences between these separations, according to the probability distribution of the transitions $e \rightleftarrows f$ and $e \rightarrow e$, $f \rightarrow f$.

The lesson of this for present purposes is that if one wishes to get beyond a first approximation in the treatment of the energies of the fine-structure levels of $2p\,^3\Pi$ the effects of the fine structure of the upper states must be eliminated, as these may, possibly, be comparable with the *changes* in the $2p\,^3\Pi$ states which are under investigation.

6. THE VARIATION OF $\Delta\nu$ WITH K''

The magnitude of the doublet width of any of the lines under consideration can be put in the form $F''_f - F''_e + \alpha(F'_f - F'_e)$, where α is a function of the transition probabilities from the fine-structure levels of the upper to those of the lower state. α may be either positive or negative. We assumed that it is independent, to a sufficient approximation, of the difference of the quantum numbers of the final states F'' to which the expression is applied. Otherwise, the following argument requires reconsideration.

In this notation the doublet width of any $R(K)$ line, for example, may be written

$$\Delta\nu R(K) = F''_f(\nu'', K) - F''_e(\nu'', K) + \alpha(F'_f(\nu', K+1) - F'_e(\nu', K+1)),$$

then

$$\Delta\nu P(K+2) = F''_f(\nu'', K+2) - F''_e(\nu'', K+2) + \alpha(F'_f(\nu', K+1) - F'_e(\nu', K+1))$$

and

$$\Delta\nu R(K) - \Delta\nu(P(K+2)) = F''_f(\nu'', K) - F''_e(\nu'', K) - (F''_f(\nu'', K+2) + F''_e(\nu'', K+2)),$$

is the change in the width of fine-structure levels of the final state at $\nu = \nu''$ as K changes to $K-2$. There is not a great wealth of material to apply this to. What there is is set out in table 11. The regular $R1 - P3$ values which are almost complete show a steady falling off of the separation as ν'' increases from 0 to 3, the mean values, for what should be the same quantities, being: at $0 \rightarrow 0$, $+0{\cdot}039$, at $1 \rightarrow 1$, $+0{\cdot}038_3$, at $2 \rightarrow 2$, $+0{\cdot}028$; and at $3 \rightarrow 3$, $+0{\cdot}011_5$. The values (one each) of $R'2 - P'4$ and $R3 - P5$ show that the $R(K) - P(K+2)$ differences increase with increasing K''.

The figures for the irregular levels point to a much greater separation of the fine-structure levels at $R2P4$ than at $R'1P'3$, and suggest that the increase of separation with increasing K'' extends to these levels, that is to say to $2p\,^3\Pi_d$, also.

The data are not very different from what would be obtained if the energy of the levels giving rise to the fine structure consisted of two parts, one proportional to the rotational energy of the molecule and the other independent of it. The rotational energy will be roughly proportional to K^2. If this were exact, the energy differences for $R1$, $P3:R'2$, $P'4:R3$, $P5$ would be as $2:3:4$. For $3'$, $3''$, $R1P3 = +0{\cdot}011_5$, which would give a computed value for $3'$, $3''$, $R'2P'4$ of $+0{\cdot}017_3$ as against the observed ? $+0{\cdot}026$ and from $1'$, $1''$ $R1$, $P3 = +0{\cdot}038_5$ we calculate

$$1', 1''\ R3 - P5 = +0{\cdot}077$$

as against ? $+0{\cdot}072$ observed. These agreements are as close as the experimental results would justify expectation, and the data would hardly warrant an attempt to refine the theory further. Frequency differences such as those of $R1 - P3$ eliminate the effects of differences in the energies of the fine-structure levels which do not depend on K and so disclose nothing about their magnitude.

Something may be said about this if one turns back to the doublet separations $\Delta\nu$ and forget about their differences for the time being. Actually these are greatest for the smallest K values (table 10) and diminish as K increases. If the part of the energy difference of the fine-structure levels of the upper level which is independent of K is denoted by a' and the corresponding quantity for the lower level by a'', $b'(K')$ and $b''(K'')$ similarly denoting the parts which depend on K, then

$$\begin{aligned}
\Delta\nu(R(K)) &= a' + b'(K+1) + a'' + b''(K),\\
\Delta\nu(P(K)) &= a' + b'(K-1) + a'' + b''(K),\\
\Delta\nu(R(K)) - \Delta\nu(P(K)) &= b'(K+1) - b'(K-1),\\
\Delta\nu(R(K-1)) &= a' + b'(K) + a'' + b''(K-1),\\
\Delta\nu(P(K+1)) &= a' + b'(K) + a'' + b''(K+1),\\
\Delta\nu(R(K-1)) - \Delta\nu(P(K+1)) &= b''(K-1) - b''(K+1).
\end{aligned}$$

An examination of the data shows that the $b'(K+1) - b'(K-1)$ are mostly negative, i.e. $b'(K) > b'(K+2)$ and the $b''(K-1) - b''(K-1)$ are mostly positive, i.e. $b''(K) > b''(K+2)$. This at any rate seems to be what happens at ν', $\nu'' = 0$ or 1 where the data are most plentiful. At higher ν'' the diminution due to the increasing numerical value of the b term seems to be counteracted by something else, and in the irregular group there is no clear sign of any falling off of the doublet separation with increasing K in the $3 \rightarrow 3$ bands.

It thus appears that the change with K for the upper levels is on the whole in the same sense as that for the $2p\,^3\Pi$ levels. We hope to deal with this more fully in a later paper.

If the irregular levels of the $3 \rightarrow 3$ band be excluded, then it is found that the total doublet separation is greatest at the lowest K'' level ($K'' = 1$) and then diminishes as K'' increases, slowly at first and then more rapidly. The part $a = a' + a''$ of the doublet width which is independent of K is the extrapolated width at $K'' = 0$, and is thus a little greater in magnitude than the width at $K'' = 1$. At $K'' = 0$, $a' + a''$ is the whole width and at $K'' = 1$ it is about 95 % of the whole. There is evidence that $a' + b'$ (1) does not exceed about 5 % of the whole, so that a' is almost certainly within a few percent of the average width at $K'' = 1$ in table 10. Since the numerical value of both $b''(K)$ and $b'(K)$ increase as K increases and the total width diminishes as K increases, it follows that $b''(K)$ (and $b'(K)$) have the opposite sign to a''. It appears also that at $K = 5$, $b''(K)$ is about $\frac{1}{2}a''$.

Comparison with the helium band spectrum

The only band spectrum to exhibit a fine structure similar to that of H_2 is the spectrum of He_2. This He_2 fine structure was discovered by Monk & Mulliken in 1929. The resemblance between the two is very close indeed. The most obvious differences are due to the fact that in the He_2 spectrum the alternate dashed lines in the H_2 spectrum are missing owing to the zero value of the spin of the helium nucleus. Thus in He_2 the number of band lines is half the number in H_2 for the same K range, and there is nothing to disclose the behaviour of the dashed group of lines. The observations and measurements for He_2 were all made on $0 \rightarrow 0$ bands.

As in H_2 the only lines in He_2 for which the doublets are observed as clearly resolved are those which have $2p\,^3\Pi$ as the lower state. These fall into two groups, denoted by A and B, corresponding to our regular and irregular groups. In all cases the component of higher frequency is the more intense. For the A (regular) group the average doublet distance at $v'' = 0$ falls from 0·36 at $K = 1$ through 0·27 at $K = 5$ to 0·22 at $K = 9$. This corresponds to 0·204 at $K = 1$ through 0·168 at $K = 3$ to about 0·15 at $K = 5$ for H_2 at $v'' = 0$ (table 10). The value for He_2 starts by being nearly twice that for H_2 at $K = 1$, and they both fall off in about the same relative proportions as K increases. For these doublets, Mulliken & Monk (1929) say the intensity ratio was about the same, about 2:1 to 3:1 as estimated visually. Our mean for H_2 at $v'' = 0$ is $2{\cdot}4_5$ (table 10). For the B (irregular) levels in He_2 only the lowest ($K'' = 2$) levels were resolved, the higher $K'' = 4$, 6 and 8 levels appearing as symmetrically broadened lines. At the $K'' = 2$ level the mean $\Delta\nu$ for He_2 was found to be 0·21 cm. Our mean values for H_2 at $v'' = 0$ fell from 0·15 at $K'' = 1$ to $0{\cdot}07_4$ at $K'' = 4$ (table 10) so that the splitting in H_2 is probably closely parallel to that in He_2 on about half the scale. The rapid contraction of the doublets at $v'' = 0$ is not observed in H_2 at higher v''. For He_2 the intensity ratio of the resolved $K'' = 2$ doublets is given as from 3:1 to 6:1. Our measures of the intensity ratio of the corresponding doublet for H_2 at $v'' = 0$ give the weighted mean value $4{\cdot}0_7$.

Mulliken & Monk give a valuable discussion of the theory of these fine structures which is particularly interesting where it concerns the intensities of the components (Monk & Mulliken 1929; Mulliken & Monk 1929). According to this, which is based on a theory of Kramers (1929), the energy differences depend entirely on the interaction energy of the spins s_1 and s_2 of the two outer electrons, which in the triplet states combine to form the spin quantum number of the molecule $S = s_1 + s_2 = 1$. For each K the associated S can take the values 0 ± 1 giving three close triplet levels with $J = K$, $K \pm 1$. Since the weights of the states are $= 2J + 1$ the intensities of the components will be in the ratio $2K + 3 : 2K + 1 : 2K - 1$. In Σ states, according to Kramers' theory, the states with $J = K \pm 1$ have almost identical energy, and this is attributed by Mulliken & Monk to the $2\Pi_c$ (A or regular) states. They are supposed to be so close together as to be completely unresolved and to be lower than the K level, thus giving the observed doublet. The ratio of the intensities of the combined lower levels to the upper level is $(2K + 3 + 2K - 1) \div (2K + 1) = 2$ for all K. For the $2\Pi_d$ (B or irregular) levels which are resolved into doublets the $J = K$, $K + 1$ sublevels are so close as to be unresolved and the $J = K - 1$ level lies above them. For the ratio of the intensities of the components of the resolved doublets, all of which have $K'' = 2$, should therefore be $7 + 5 \div 3 = 4$. All these intensity predictions are in agreement with Monk & Mulliken's observations on the He_2 lines within the limits of accuracy they seem to claim.

Remembering the very close similarity between the fine-structure results of the He_2 and H_2 spectra and in view of the striking successes of these intensity relations in He_2 one is tempted to carry the theory over bodily from He_2 to H_2. At first sight this seems to be crowned with success, for our H_2 intensities make $I(a)/I(b)$ for the regular group almost independent of all the variables and closer to 2 than to 3, and the value of $I(a)/I(b)$ for the irregular group at $K'' = 2$ is $4{\cdot}2_8$ when averaged over ν'' for 0–3 and actually $4{\cdot}0_7$ at $\nu'' = 0$ against the predicted 4. However, a closer inspection does not place the matter in quite such a rosy light.

In the first place, the discrepancy between about 2·4, our value of $I(a/b)$ for the A group, exceeds 2 by more than we should have expected. It is true that there are great difficulties in measuring these intensities and their ratios exactly. It may be that our methods are exaggerating the $I(a/b)$ ratio in favour of the stronger component.

A more serious difficulty arises at the B levels. If the ratio is 4 at the $K = 2$ level, on the same basis at the $K = 1$ level it should be $(2K + 3 + 2K + 1)/(2K - 1) = 8$. Our highest single measured photometric value for the irregular levels is 6·1. This is for the line 16802·37 (10G) $3s\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$, $1 \rightarrow 1$, $P2$. This particular line is difficult to measure accurately on our plates owing to a very heavy background. Our final assessment of $I(a/b)$ for this line, taking into account another (lower) photometric measure and some visual estimates, is 4 to 6 (4); the weight (4) means that a value between 4 and 6 is fairly certain, it does not mean that 6 is more likely than 4, but it does imply that the most probable value is close to 5. For the B levels at $K'' = 1$ we have no photomicrometric measures. These are difficult owing to the weakness of the b components of the composite lines none of which have much intensity. Of the eleven eye estimates for $R'1$, $Q'1$, $P'1$ lines of the $3s\,{}^3\Sigma \rightarrow 2p\,{}^3\Pi$ and the $3d\,{}^3\Sigma$, ${}^3\Pi$, ${}^3\Delta$ systems the three highest have $I(a/b) \geqslant 5$ (0), $\geqslant 5$ (0) and ?6 (5). None of the other eight exceed 5 and the weighted mean of all the eleven we put at 5·4. A case could be made out for its being closer to 5 even than 5·4.

It seems very improbable that $I(a/b)$ for the $R'1$ $Q'1$ $P'1$ lines can be as high as 8. It is true that there is not a photometric confirmation at this level, but this method has been checked against visual estimation for a number of undashed lines ending on each of the levels $K'' = 1$, 2 and 4 and on the dashed lines ending on $K'' = 2$. There is no drastic change in the conditions which would warrant any expectation of an error approaching a factor of 2 in changing from these to the dashed lines $R'1$, $Q'1$, $P'1$.

This co-ordination of the results, both for H_2 and He_2, could be made consistent by increasing all the J values by $\frac{1}{2}$. This would not affect the calculated value 2 of $I(a/b)$ for the A levels, since the ratio $2(K+1+\frac{1}{2})+1+2(K-1+\frac{1}{2})+1/2(K+\frac{1}{2})+1$ is equal to 2 whatever quantity may be added to K or J. On the other hand, the calculated $I(a/b)$ for the B levels $(2(K+1+\frac{1}{2})+1+2\ (K+\frac{1}{2})+1)/2(K-1+\frac{1}{2})+1)$ becomes, for each K,

K	1	2	3	4
$I(a/b)$ calc.	5	3·5	3	2·75
$I(a/b)$ exp. H_2	5·4	$4{\cdot}2_8$	3·7	$0{\cdot}9_4$

At all K except $K = 4$ the observed values for H_2 are all a little higher than the calculated ratios just as was the case for the A values. The sharp drop to about 1·0 at $K = 4$, which seems to be real enough, may be due to a relative reorientation of the sublevels similar to that postulated by Monk & Mulliken to account for the contraction of the $\Delta\nu$'s at higher K'' values in He_2 at $\nu'' = 0$. This also sets in at $K'' \geqslant 4$.

No doubt the close similarity of both spectra means that the explanation of their fine spectra is fundamentally the same for both He_2 and H_2. For each, a considerable body of facts has been assembled but not enough to cover all the possibilities. Further consideration of these seem to be premature and we shall postpone it for the present. An opportunity to return to it may perhaps occur in later papers in which we hope to consider the possibility of fine structure in levels of H_2 states other than $2p\,{}^3\Pi_{cd}$.

We wish to express our thanks to the Department of Scientific and Industrial Research and to the Royal Society for the provision of financial aid which enabled this work to be carried out.

REFERENCES

Gale, H. G., Monk, G. S. & Lee, K. O. 1928 *Astrophys. J.* **67, 89.**
Kramers, H. A. 1929 *Z. Phys.* **53**, 422.
Kapuscinsky, W. & Eymers, J. G. 1929 *Proc. Roy. Soc.* A, **122**, 58.
Monk, G. S. & Mulliken, R. S. 1929 *Nature*, **124**, 91.
Mulliken, R. S. & Monk, G. S. 1929 *Phys. Rev.* **34**, 1530.
Richardson, O. W. 1929 *Trans. Faraday. Soc.* **25**, 691.
Richardson, O. W. 1930 *Proc. Roy. Soc.* A, **126**, 501.
Richardson, O. W. 1934 *Molecular hydrogen and its spectrum.* New Haven: Yale University Press; London: Oxford University Press.

Richardson, O. W. & Davidson, P. M. 1931 *Proc. Roy. Soc.* A, **131**, 658.
Richardson, O. W. & Davidson, P. M. 1933 *Proc. Roy. Soc.* A, **140**, 25.
Richardson, O. W., Davidson, P. M., Marsden, J. & Evans, W. M. 1933 *Proc. Roy. Soc.* A, **142**, 40.
Richardson, O. W. & Rymer, T. B. 1934*a* *Proc. Roy. Soc.* A, **147**, 24.
Richardson, O. W. & Rymer, T. B. 1934*b* *Proc. Roy. Soc.* A, **147**, 251.
Richardson, O. W. & Williams, W. E. 1931 *Nature*, **127**, 481.
Sandeman, A. 1932 *Proc. Roy. Soc.* A, **138**, 395.

Radiative equilibrium in the ionosphere

By R. v. d. R. Woolley

(*Communicated by S. Chapman, F.R.S.—Received* 20 *April* 1945)

The equations of radiative equilibrium in the earth's atmosphere are examined with special reference to high temperatures in the ionosphere arising out of a radiation balance between heavy ultra-violet absorption in the ultra-violet and relatively weak infra-red emission. It is suggested that while molecular oxygen is the principal ultra-violet absorption agent at heights below 250 km., at greater heights the absorption is mainly due to atomic oxygen. Similarly, it is suggested that while water vapour is the principal infra-red radiator at a height of 100 km., at much greater heights water vapour is absent and the temperature is effectively controlled by negative ions at a height of 250 km. and perhaps by dust at much greater heights.

The gravitational equilibrium of a dissociating oxygen atmosphere is discussed, reference being made to the effects of collisions and of absorption of the dissociating radiation.

1. The generally accepted theory that the upper layers of the earth's atmosphere are in radiative equilibrium was put forward by Gold (1909). Gold's theory was discussed by Milne (1922) in a paper to which detailed reference will be made later. In the same year Lindemann & Dobson (1922) produced observational evidence to show that the temperature increased considerably above the stratosphere. Gowan (1928) worked out a heat balance between absorption of ultra-violet sunlight by ozone and emission of temperature radiation by water vapour, and showed that temperatures of order 300° C were to be expected. In a later paper, Gowan (1930) gave temperatures between 250 and 500° C at a height of 80 km., depending on the relative amounts of these two constituents of the atmosphere. A few years later it was suggested by various authors that at greater heights the temperature attained much greater values. Thus Angenheister (1932) studied the heights of polar aurorae in summer and winter and concluded that if the winter temperature from 100 to 200 km. was 300° C, the summer temperature was 1200° C. Appleton & Naismith (1935) made a similar deduction from seasonal variations in the critical frequency of the F_2 region. Martyn & Pulley (1936) found that relative collision frequencies in the E and F_2 regions could only be explained by a temperature of order 1000° C in the atmosphere between these regions, and put forward the hypothesis that this high

temperature persisted all the year round. Shortly afterwards Godfrey & Price (1937) found theoretical temperatures of order 1000° C by working out a heat balance between infra-red water vapour emission and ultra-violet absorption by molecular oxygen (as well as ozone). These authors followed Gowan's method, which considers a single layer, without working out the temperature, as a function of height from the equations of transfer of radiation, as was done by Milne.

The present paper develops equations for the temperature as a function of the height in a form suitable for distinguishing between the various molecular processes of re-emission of light following upon absorption. It is pointed out that dissociation of oxygen affects the heat balance profoundly, and expressions for the dissociation of molecular oxygen as a function of height are worked out. The paper is concerned with heights above the ozone layer and a discussion of the formation of ozone is explicitly avoided.

2. In any atmosphere, let $I_\nu(\theta)\,d\nu\,d\omega/4\pi$ be the flow of radiation per sq.cm./sec. in frequencies between ν and $\nu+d\nu$ in directions falling within a cone of solid angle $d\omega$ inclined at an angle θ to the vertical, which is the axis of x taken positive upwards. Consider the passage of this radiation through a cylinder of unit cross-section and of length $ds = dx \sec\theta$. The radiation will lose by absorption the amount

$$k_\nu I_\nu(\theta)\,\rho\,ds\,d\nu\,d\omega/4\pi,$$

where k_ν is the coefficient of absorption per unit mass of the atmosphere. The energy which is removed in this way may be merely scattered, or it may be converted into kinetic energy of the particles of the gas, or it may be emitted in frequencies outside the range from ν to $\nu+d\nu$ by fluorescence. Let the fraction ξ_ν of the energy absorbed be scattered, the fraction η_ν be converted into kinetic energy, and let the fraction ζ_ν be transferred to other frequencies, where $\xi_\nu+\eta_\nu+\zeta_\nu = 1$.

Now examine the emission with the cone $d\omega$ in frequencies within the range ν to $\nu+d\nu$ by the cylinder. If the scattered radiation is isotropic, the amount scattered into $d\omega$ is

$$\xi_\nu k_\nu \rho\,ds\,d\nu\,d\omega/4\pi \times J_\nu,$$

where J_ν is the average value of $I_\nu(\theta)$ over a sphere, or $4\pi J_\nu = \int I_\nu(\theta)\,d\omega$. The absorption $\eta_\nu k_\nu$ is balanced by thermal emission. The balance is exact in thermodynamical equilibrium, and the amount emitted depends only on the kinetic energy of the molecules of the gas. If the molecules have a Maxwellian velocity distribution appropriate to some temperature T, then the emission is

$$\eta_\nu k_\nu \rho\,ds\,d\nu\,d\omega/4\pi \times B(\nu, T),$$

where $B(\nu, T)$ is the intensity of equilibrium (black-body) radiation,

$$= 8\pi h\nu^3 c^{-2}(e^{h\nu/RT}-1)^{-1}.$$

The symbols h, R stand for Planck's constant and Boltzmann's constant.

Finally, there will be some emission into the frequency range ν to $\nu+d\nu$ by fluorescent transfer of energy from other frequencies. Let this be

$$\zeta_\nu k_\nu \rho\, ds\, d\nu\, d\omega/4\pi \times E_\nu.$$

Summing up the gains and losses to the flow $I_\nu(\theta)\, d\nu\, d\omega/4\pi$ on account of its passage through the cylinder, then

$$\cos\theta \frac{dI_\nu(\theta)}{k_\nu \rho\, dx} = -I_\nu(\theta) + \xi_\nu J_\nu + \eta_\nu B(\nu, T) + \zeta_\nu E_\nu. \tag{2·1}$$

Now define two further quantities H_ν and K_ν by

$$4\pi H_\nu = \int I_\nu(\theta) \cos\theta\, d\omega, \quad 4\pi K_\nu = \int I_\nu(\theta) \cos^2\theta\, d\omega,$$

both integrals being taken over a sphere. H_ν is the flux of radiation in the direction x and K_ν may be thought of as the potential of this flux. Then averaging equation (2·1) over a sphere,

$$\frac{dH_\nu}{k_\nu \rho\, dx} = -J_\nu(1-\xi_\nu) + \eta_\nu B(\nu, T) + \zeta_\nu E_\nu. \tag{2·2}$$

Multiplying equation (2·2) by $k_\nu d\nu$ and integrating, then

$$\frac{dH}{\rho\, dx} = -\int (1-\xi_\nu) k_\nu J_\nu d\nu + \int \eta_\nu k_\nu B(\nu, T)\, d\nu + \int \zeta_\nu k_\nu E_\nu d\nu,$$

writing H for $\int H_\nu d\nu$. If there are no sources or sinks of energy in the atmosphere, the net flux is constant (although H_ν is not so), and hence

$$\int (1-\xi_\nu) k_\nu J_\nu d\nu = \int \eta_\nu k_\nu B(\nu, T)\, d\nu + \int \zeta_\nu k_\nu E_\nu d\nu.$$

Since fluorescence only transfers energy from one frequency to another

$$\int \zeta_\nu k_\nu E_\nu d\nu = \int \zeta_\nu k_\nu J_\nu d\nu.$$

Hence, when there are no sources or sinks,

$$\int \eta_\nu k_\nu J_\nu d\nu = \int \eta_\nu k_\nu B(\nu, T)\, d\nu. \tag{2·3}$$

This is the equation of radiative equilibrium which determines the temperature T of the atmosphere as judged by the kinetic energy of the particles. In the remainder of this paper it will be assumed that the velocities of the particles are distributed according to Maxwell's law, and the parameter T in that law will be referred to as the temperature of the atmosphere.

Multiplying equation (2·1) by $\cos\theta$ and integrating over a sphere, then

$$\frac{dK_\nu}{k_\nu \rho\, dx} = -H_\nu. \tag{2·4}$$

(It may throw some light on the nature and importance of the quantity $K = \int K_\nu d\nu$ to remark that $\nabla^2 K = -k_\nu \rho \operatorname{div} H + \int \rho \frac{dK_\nu}{d\nu} \operatorname{div} H\, d\nu$, so that if there are no sources or sinks of energy and $\operatorname{div} H = 0$, then $\nabla^2 K = 0$. In most applications, however, there is either plane or spherical symmetry and the equation reduces to simpler forms.)

Advance has been made in the treatment of radiative equilibrium in the solar atmosphere by assuming a relation between K_ν and J_ν based on some simple assumption about the distribution of $I_\nu(\theta)$ with θ. For example, if $I_\nu(\theta) = I_1$ when $0 < \theta < \frac{1}{2}\pi$, and if $I_\nu(\theta) = I_2$ when $\frac{1}{2}\pi < \theta < \pi$, then $K_\nu = \frac{1}{3}J_\nu$. Again, if $I_\nu(\theta) = a + b\cos\theta$, once more $K_\nu = \frac{1}{3}J_\nu$. The solution of the simplest astrophysical problem is as follows. Consider an atmosphere in which there are no absorption lines, so that $\xi_\nu = 0$ and ignore fluorescence, so that $\zeta_\nu = 0$. Then $\eta = 1$. Further, suppose that k_ν is independent of ν and equal to k and $d\tau = -k\rho\, dx$. Then using $K = \frac{1}{3}J$ it follows that

$$J = 3H\tau + \text{constant},$$

the boundary condition being no inflow of radiation at the surface of the star or $\int I_\nu(\theta)\, d\omega = 0$ over a hemisphere equal to zero where $\tau = 0$. This gives $J = 2H$ at $\tau = 0$ according to one approximation ($I_\nu(\theta)$ constant over a hemisphere) and $J = \frac{4}{3}H$ according to the other ($I_\nu(\theta) = a + b\cos\theta$). Adopting the former, then

$$J = H(2 + 3\tau).$$

This gives the temperature, since $J = B(T) = acT^4$ and

$$T^4 = T_0^4(1 + \tfrac{3}{2}\tau).$$

This equation is due to Milne. It receives support from the observed darkening of the solar limb.

The approximation $K_\nu = \frac{1}{3}J_\nu$ has had further successes in accounting for the main features of absorption lines, and further mathematical approximation has shown that in some cases at any rate the errors involved in assuming that $K_\nu = \frac{1}{3}J_\nu$ are small.

3. It may be asked whether this approximation is of any value in dealing with the earth's atmosphere. In this case it is convenient to assume that the flow of radiation has the intensity $I'_\nu(\theta)$ in all directions except in a cone of solid angle $4\pi\omega$ (the cone subtended by the sun's disk) in which the intensity is $I'_\nu(\pi - z) + S_\nu$ where z is the sun's zenith distance. Write

$$4\pi J'_\nu = \int I'_\nu(\theta)\, d\omega, \quad 4\pi H'_\nu = \int I'_\nu(\theta)\cos\theta\, d\omega, \quad 4\pi K'_\nu = \int I'_\nu(\theta)\cos^2\theta\, d\omega,$$

all the integrals being taken over a sphere. Then

$$J_\nu = J'_\nu + \omega S_\nu, \quad H_\nu = H'_\nu - \omega S_\nu \cos z, \quad K_\nu = K'_\nu + \omega S_\nu \cos^2 z.$$

The radiation outside the special cone may be supposed evenly distributed according to some simple law such as $I'_\nu(\theta) = a + b\cos\theta$, so that it may be supposed that $K'_\nu = \frac{1}{3}J_\nu$. It is at once clear that in the special case $\sec^2 z = 3$, $K_\nu = \frac{1}{3}J_\nu$ also. It is also clear that $K_\nu = \cos^2 z J_\nu$ will be a better approximation than $K_\nu = \frac{1}{3}J_\nu$ if direct sunlight preponderates over a scattered sunlight.

In what follows the intensity $I_\nu(\theta)$ will be used, defined by

$$\begin{aligned} I_\nu(\theta) &= I'_\nu(\theta) && \text{outside the special cone } 4\pi\omega, \\ I_\nu(\theta) &= I'_\nu(\pi - z) + S_\nu && \text{inside the special cone } 4\pi\omega. \end{aligned}$$

A distinction of this kind was made by Milne (1922).

By definition, $4\pi\omega$ is the solid angle subtended by the solar disk. According to Auwers the mean solar diameter is 31′ 59″, which gives $\omega = 5{\cdot}410 \times 10^{-6}$. Let ρ be the radius of the sun, and let R be the radius of the earth's orbit. Let Σ_ν be the emission per sec./sq.cm. of the sun's surface. The total emission from the sun is then $4\pi\rho^2\Sigma_\nu$. Since all this radiation passes through a sphere of radius R the amount which passes through 1 sq.cm. normal to the sun at the earth's mean distance per sec. is $\rho^2/R^2 . \Sigma_\nu$. But this is $4\pi\omega S^0_\nu$, where S^0_ν is the value of S_ν at the top of the earth's atmosphere. Again $4\pi\omega = \pi\rho^2/R^2$, so that

$$\pi S^0_\nu = \Sigma_\nu.$$

The value of Σ_ν is not known in the far ultra-violet regions of the spectrum in which present interest lies. In the visible spectrum and in the infra-red, however, it is known that the distribution in frequency of solar radiation is close to that of a black body at 6850° (Plaskett), while the effective temperature is known, more reliably, to be 5740°. If a colour temperature of 6825° be assumed ($= 5740° \times 2^{\frac{1}{4}}$),

then
$$\Sigma_\nu = \frac{\pi h\nu^3}{c^2}(e^{h\nu/RT'} - 1)^{-1},$$

where $T' = 6825°$. This gives $\int \Sigma_\nu d\nu = \frac{1}{8}acT'^4 = \frac{1}{4}acT_e^4$ with $T_e = 5740$. Now

$$S^0_\nu = h\nu^3 c^{-2}(e^{h\nu/RT'} - 1)^{-1},$$

and since for ultra-violet radiation $e^{h\nu/RT'}$ is much greater than 1, then

$$S^0_\nu = h\nu^3 c^{-2} e^{-\beta\nu}, \tag{3·1}$$

where $\beta = 6{\cdot}991 \times 10^{-15}$ and $\log\beta = \overline{15}.8446$. This value of β is applicable to the visible spectrum. As it is likely that the solar absorption coefficient increases with increasing frequency, it is probable that in the far ultra-violet the solar radiation is less than is given by (3·1) with $T' = 6825°$. In particular, it is very likely that there is a drop in the continuous spectrum at the head of the Lyman series at 912 A.

4. In this section an atmosphere is considered in which there is neither scattering nor fluorescence, that is to say, $\xi = \zeta = 0$ and $\eta = 1$. The case in which k_ν is independent of ν was treated by Milne. It has no application to the earth's upper atmosphere in which k_ν has large values in the infra-red and ultra-violet and is practically zero in the visible spectrum. Now pass on to a mathematical case which, though far too crude to give an accurate representation of the actual atmosphere, illustrates several points, and in particular the reversal of the temperature gradient. Let $k_\nu = k_1$ from ν_1 to $\nu_1 + \Delta\nu_1$ (infra-red frequencies), and let $k_\nu = k_2$ from ν_2 to $\nu_2 + \Delta\nu_2$ (ultra-violet frequencies), and let k_ν be zero for all other frequencies. Then from (2·1)

$$\cos\theta \frac{dI_1(\theta)}{\rho\, dx} = -k_1 I_1(\theta) + k_1 B(\nu_1, T),$$

$$\cos\theta \frac{dI_2(\theta)}{\rho\, dx} = -k_2 I_2(\theta) + k_2 B(\nu_2, T).$$

In the earth's atmosphere T will be so small that $B(\nu_2, T)$ may be neglected in comparison with $\omega S^0_{\nu_2}$ (in the ultra-violet frequencies). In this case

$$\frac{dH_2}{\rho\, dx} = -k_2 J_2, \quad \frac{dK_2}{\rho\, dx} = -k_2 H_2.$$

The ultra-violet light is almost entirely direct sunlight so that $K_2 = \cos^2 z J_2$, and the equations give $\cos^2 z \dfrac{d^2 J_2}{\rho^2 dx^2} = k_2^2 J_2$, or

$$J_2 = J_2^0 \exp - \left\{ k_2 \sec z \int_x^\infty \rho\, dx \right\},$$

which is ordinary exponential absorption. Now in radiative equilibrium with no sources or sinks of energy it follows from (2·3) that

$$k_1 \Delta\nu_1 \{J_1 - B(\nu_1, T)\} + k_2 \Delta\nu_2 J_2 = 0.$$

Write $\Delta\nu_2 = \lambda \Delta\nu_1$ and $k_2 = \alpha k_1$ and let $d\tau = -k_1 \rho\, dx$. Then

$$\frac{dH_1}{d\tau} = J_1 - B(\nu_1, T) = -\alpha\lambda J_2^0 e^{-\alpha\tau \sec z},$$

$$H_1 = A + \lambda \cos z\, J_2^0 e^{-\alpha\tau \sec z}.$$

Notice that $H = H_1 \Delta\nu_1 + H_2 \Delta\nu_2$ is constant, as it must be if there are no sources or sinks.

In the infra-red, temperature radiation is more important than direct sunlight and $K_1 = \frac{1}{3} J_1$. Then

$$dJ_1/d\tau = 3H_1 = 3A + 3\lambda \cos z\, J_2^0 e^{-\alpha\tau \sec z},$$

which gives $J_1 = J_1^0 + 3A\tau + 3\alpha^{-1}\lambda \cos^2 z J_2^0 + J_2^0 e^{-\alpha\tau \sec z}(\alpha\lambda - 3\alpha^{-1}\lambda \cos^2 z).$

But $$B(\nu_1, T) = J_1 + \alpha\lambda J_2.$$

Hence $$\frac{dB(\nu_1, T)}{d\tau} = 3A + \lambda \sec z\, J_2^0 e^{-\alpha\tau \sec z}(3 \cos^2 z - \alpha^2).$$

Now at the lower boundary of the radiative region (the tropopause), which is taken as $\tau = \tau_1$, infra-red radiation is given out by the earth so that H_1 has a positive value, say $\mathscr{H}$. Then

$$\mathscr{H} = A + \lambda \cos z J_2^0 e^{-\alpha\tau_1 \sec z},$$

it may be supposed that no ultra-violet radiation reaches the lower boundary and $J_2^0 e^{-\alpha\tau_1 \sec z}$ may be neglected. This gives $A = \mathscr{H}$, and accordingly

$$\frac{dB(\nu_1, T)}{d\tau} = 3\mathscr{H} - \lambda \sec z J_2^0 (\alpha^2 - 3\cos^2 z)\, e^{-\alpha\tau \sec z}.$$

If α is large $3\cos^2 z$ may be neglected compared with α^2. Then the temperature gradient changes sign at $\tau = \tau'$, where

$$\mathscr{H} = \tfrac{1}{3}\lambda\alpha^2 J_2^0 \sec z\, e^{-\alpha\tau' \sec z}.$$

The boundary condition at the top of the atmosphere is no inflow of infra-red radiation, or $J_1 = 2H_1$. (The solar infra-red radiation has in intensity of order $\omega T'/T$ compared with the earth's radiation.) Then $J_1^0 = 2(\mathscr{H} + \lambda \cos z J_2^0)$. But $B(\nu_1, T) = J_1 + \alpha\lambda J_2$ so that if $T = T_0$ at $\tau = 0$

$$B(\nu_1, T_0) = 2\mathscr{H} + \lambda J_2^0 (\alpha + 2\cos z).$$

The temperature at the point where the gradient is reversed $\tau = \tau'$ is $T_{\text{min.}}$, where

$$B(\nu_1, T_{\text{min.}}) = 2\mathscr{H}(1 + \tfrac{3}{2}\tau' + 3\alpha^{-1}\sec z) + \lambda J_2^0 (2\cos z + 3\alpha^{-1}\cos^2 z).$$

For α large these expressions become

$$B(\nu_1, T_0) = 2\mathscr{H} + \alpha\lambda J_2^0, \tag{4·1}$$

$$B(\nu_1, T_{\text{min.}}) = 2\mathscr{H}(1 + \tfrac{3}{2}\tau') + 2\lambda \cos^2 z J_2^0, \tag{4·2}$$

giving the temperature at the top of the atmosphere T_0 and the minimum temperature $T_{\text{min.}}$. $\mathscr{H}$ is supposed known and $J_2^0 = \omega S_2^0$. Equation (4·1) is very similar to Milne's equation (28″), from which it differs slightly because Milne's approximation is not quite the same as $K = \frac{1}{3}J$, and also because an internal boundary has been used in deriving (4·1). The use of this internal boundary shows that a minimum temperature can occur in an atmosphere in radiative equilibrium. It should be noticed that Milne's statement that $\cos\alpha > \frac{1}{2}n$ (in our notation $\cos z > \sqrt{3}\,\alpha$), in most applications, is correct if one is comparing heavy infra-red absorption with an almost transparent visible spectrum; but the introduction of still heavier absorption in the far ultra-violet leads to α large with, of course, very different consequences. Milne's use of exponential absorption for direct sunlight and his use of (24) and (25) for earth radiation correspond to our approximation $K_2 = \cos^2 z J_2$ and $K_1 = \frac{1}{3}J_1$.

5. One cannot proceed from (4·1) to an evaluation of the actual temperatures which will be encountered in the atmosphere because k_ν is not even roughly independent of ν in the far ultra-violet. Consider first the ultra-violet absorption due to molecular oxygen O_2. The absorption is in the Schumann-Runge continuum and is

absorbed by the ground state $x^3\Sigma$. It results in dissociation of the molecule into two atoms, one in the ground state ^{3}P and one in a metastable state ^{1}D, thus

$$O_2(x^3\Sigma) + h\nu \rightarrow O(^3P) + O(^1D).$$

The dissociation potential of O_2 is 5·09 V and the excitation potential of $O(^1D)$ is 1·96 V. Hence the excess of energy over 7·05 V (corresponding to $\nu = 5{\cdot}7 \times 10^4$ cm.$^{-1}$) absorbed in the Schumann-Runge continuum is converted into kinetic energy. Turning to the potential energy, the most likely transitions for the $O(^1D)$ atom are:

I. Ordinary recombination $O(^1D) + O(^3P) \rightarrow O_2 + h\nu$, contributing to ξ.

II. (*a*) Super-elastic collision $O(^1D) + X \rightarrow O(^3P) + X + \frac{1}{2}mv^2$, contributing to η.

(*b*) Super-elastic collision $O(^1D) + X \rightarrow O(^3P) + X'$, contributing to η.
$X' \rightarrow X + h\nu$

III. Forbidden transition $O(^1D) \rightarrow O(^3P) + h\nu$, contributing to ζ.

Of these the forbidden transition is by far the most likely when the number of particles per c.c. is 10^{14} or less, as the mean life of the metastable state is only 0·5 sec. Accordingly, scattering is negligible: most of the energy absorbed is transferred to other frequencies by fluorescence and hence

$$\xi = 0, \quad \eta = (\nu - \nu_0)/\nu, \quad \zeta = \nu_0/\nu. \tag{5·1}$$

Now work out the temperature distribution in an atmosphere in which absorption is performed by molecular oxygen, leaving atomic oxygen out of account for the moment. Define a mean absorption coefficient for infra-red frequencies by

$$\bar{k}B(T) = \int k_\nu B(\nu, T)\, d\nu,$$

the integral being taken over the infra-red, and T being that of some representative level in the atmosphere. Strictly speaking $\int \eta_\nu k_\nu J_\nu d\nu$ over the infra-red frequencies is not necessarily equal to $\bar{k}J_1$, but it will be supposed that this is so: then the equation of radiative equilibrium (2·3) gives

$$\bar{k}(J_1 - B_1) + \int_{\nu_0}^{\infty} \eta_\nu k_\nu J_\nu d\nu = 0, \tag{5·2}$$

where ν_0 is the low-frequency limit of the ultra-violet absorption.

The absorption coefficient of O_2 in the Schumann-Runge continuum has been measured by Ladenburg & van Voorhis (1933), according to whom it increases almost linearly from $\nu = 5{\cdot}9 \times 10^4$ cm.$^{-1}$ to $\nu = 6{\cdot}9 \times 10^4$ cm.$^{-1}$, decreasing sharply thereafter. For present purposes take $k_\nu = a(\nu - \nu_0)$. A rather cumbrous calculation allowing for the decrease in k_ν beyond $6{\cdot}9 \times 10^4$ cm.$^{-1}$ shows that little error is introduced by this simple assumption: the physical justification is that the solar energy curve weakens very rapidly with increasing frequency.

Now the energy from sunlight in the Schumann-Runge continuum once absorbed is re-emitted in other wave-lengths so that no radiation field is built up comparable in intensity with direct sunlight. Hence $K_\nu = \sec^2 z J_\nu$ and thus

$$J_\nu = \omega S_\nu^0 \exp\{-k_\nu \sec z/\bar{k}\tau\}.$$

Then $$\int_{\nu_0}^{\infty} \eta_\nu k_\nu J_\nu d\nu = \omega h c^{-2} \int_{\nu_0}^{\infty} a(\nu-\nu_0)^2 \nu^2 \exp\{-\beta\nu - a \sec z/\bar{k}\,.\,(\nu-\nu_0)\tau\} d\nu, \tag{5·3}$$

or with sufficient accuracy

$$\begin{aligned}\int_{\nu_0}^{\infty} \eta_\nu k_\nu J_\nu d\nu &= a\omega\nu_0^{-1} S_{\nu_0}^0 \int_0^{\infty} x^2 \exp\{-\beta(1+\mu\tau)x\} dx \\ &= 2a\omega S_{\nu_0}^0/(\nu_0\beta^3(1+\mu\tau)^3).\end{aligned} \tag{5·31}$$

Here $\mu = a \sec z/(\bar{k}\beta)$. Then from (5·2) and (5·31) it follows that

$$B_1 = J_1 + C(1+\mu\tau)^{-3},$$

where $C = 2a\omega S_{\nu_0}^0/(\bar{k}\nu_0\beta^3)$.

If an analysis similar to that in § 4 is carried out, assuming again that no ultra-violet light reaches the lower boundary $\tau = \tau_1$, then

$$\left.\begin{aligned} B(T) &= \mathscr{H}(2+3\tau) + \frac{3C}{2\mu^2}\left(\frac{2\mu}{3} + 1 - (1+\mu\tau)^{-1} + \frac{2\mu^2}{3}(1+\mu\tau)^{-3}\right), \\ B(0) &= 2\mathscr{H} + C(1+\mu^{-1}). \end{aligned}\right\} \tag{5·4}$$

Or since μ is large $$B(0) = 2\mathscr{H} + C. \tag{5·41}$$

It may be remarked that, if the approximation from (5·3) to (5·31) is not made, a more elaborate series (going up to inverse fifth powers) is obtained in place of (5·4) and, with $\beta\nu_0 = 12·4$, in place of (5·41) one obtains

$$B(0) = 2\mathscr{H} + 1·56C. \tag{5·42}$$

It is convenient to express the constant C in terms of k', the absorption coefficient when $\beta(\nu - \nu_0) = 1$. Then $\beta k' = a$ and hence

$$C = \frac{k'}{\bar{k}} \frac{2\omega S_{\nu_0}^0}{\beta^2 \nu_0}. \tag{5·5}$$

To evaluate $\mathscr{H}$, let r be the radius of the earth, ρ the radius of the sun and R the mean radius of the earth's orbit. Let the total emission from the sun be $4\pi\rho^2\Sigma$ ergs/sec. Of all the radiation which leaves the sun, the fraction $\pi r^2/4\pi R^2$ is intercepted by the earth. Of this, the fraction 0·37 is reflected as visible light (according to Abbott), while the remainder is converted into longer wave-lengths and distributed over the day and night hemispheres. Hence the average infra-red emission per sq.cm. is

$$\mathscr{H} = 0·63\Sigma \times \rho^2/4R^2 = 0·63 \times \omega\Sigma.$$

Now $\Sigma = \frac{1}{4}acT_e^4$, where a is the coefficient in Stefan's law and T_e is the effective temperature of the sun or 5740°. Hence

$$\mathscr{H} = 0·63\omega \times \tfrac{1}{4}acT_e^4. \tag{5·6}$$

But $B(T) = acT^4$, so that if $B(T) = 2\mathscr{H} \times f$, where f is some factor,

$$T = T_e \times (0{\cdot}315\omega f)^{\frac{1}{4}}$$
$$= 207^\circ{\cdot}5 \times f^{\frac{1}{4}}. \tag{5·7}$$

Again, from (3·1), (5·5) and (5·6),

$$\frac{C}{2\mathscr{H}} = \frac{k'}{\bar{k}} \times \frac{4R\nu_0^2 e^{-\beta\nu_0}}{0{\cdot}315\beta ac^3 T'^3},$$

since $T'^4 = 2T_e'^4$ has been assumed. Inserting numerical values and using (5·42), it is found that

$$T = 207^\circ{\cdot}5 \times \{1 + 7{\cdot}9 \times 10^{-5} k'/\bar{k}\}^{\frac{1}{4}}. \tag{5·8}$$

If $k'/\bar{k}$ is of order 10^8, T is of order 2000°.

The infra-red absorption due to water vapour has been measured by Hettner and also by Weber and Randall. The results do not agree, and as Hettner's observations are far more extensive they must be used. The coefficient varies irregularly with the wave-length, and Hettner's results do not extend far enough into the infra-red to give reliable results for temperatures as low as 300°. A rough estimate from Hettner's results as reproduced in Brunt's *Physical and dynamical meteorology* gives $\bar{k} = 15w$ at 2000° and $\bar{k} = 30w$ at 1000°, where w is the proportion of water vapour by weight in the atmosphere. (A very rough estimate suggests $\bar{k} = 100w$ at 300°.) The value of k' as measured by Ladenburg & van Voorhis (1933) is $1{\cdot}65 \times 10^5 w'$, where w' is the proportion of molecular oxygen by weight in the atmosphere.

While the value of w' presents no difficulty apart from dissociation, the value of w is most uncertain. On the earth's surface the water-vapour pressure is about 10 mb., but at the minimum temperature in the stratosphere (about 220°) the saturation pressure is only about 0·025 mb. This occurs at about 20 km., where the pressure of oxygen is about 27 mb., so that at this level there is about one molecule of water vapour to every thousand oxygen molecules and the ratio by mass is one to two thousand. If this ratio persists upwards the boundary temperature given by (5·8) is about 1000° C.

If, however, water molecules can pass the barrier at 20 km., possibly in the form of ice crystals, the water-vapour population may build up above this rather low limit. It does not appear impossible to ascertain the total water-vapour content above the stratosphere by actual experiment in the near future, by exposing a spectrogram in an aircraft flying at a great height.* If the above estimate of the water-vapour population is too low by a factor of 10, the boundary temperature drops from 1000° C to about 500° C. (As the temperature lowers, the value of $\bar{k}$ becomes less certain.)

By (5·4), near $\tau = 0$, $T \propto (1+\mu\tau)^{-\frac{1}{4}}$. Roughly speaking the temperature at a given level is inversely proportional to the mass of molecular oxygen above that level if the ratio of number of water-vapour molecules to number of oxygen molecules is constant. The temperature drops to one-half the boundary value below 2×10^{17}

---

* (*Note added in* 1946.) Or released from a high altitude rocket.

molecules, when the sun is in the zenith. The results of later sections may be anticipated by saying that the results of this section are thought to apply roughly to heights from 100 to 200 km.

6. At very great heights the oxygen will be dissociated to atomic O which will absorb ultra-violet radiation capable of ionizing the atom in its ground state, that is to say of wave-length below 912 A, thus

$$\mathrm{O}+h\nu\rightarrow\mathrm{O}^{+}+e.$$

The excess energy over 13·54 V (the ionization potential) appears as kinetic energy of the electrons. These will rapidly distribute their velocities according to a Maxwellian law, and the recombination spectrum will be distributed in frequency according to the molecular-velocity temperature of the (earth's) atmosphere.

It may be inquired what equilibrium would result if the atmosphere consisted entirely of one element, moreover, one which exhibited only one continuous spectrum. From each quantum $h\nu$ absorbed, the excess over the ionization energy $h(\nu-\nu_0)$ is imparted to the electrons as kinetic energy. The electrons will by collisions distribute their kinetic energy according to a Maxwellian distribution characterized by some parameter T. The emission spectrum arising from the capture of electrons with this velocity distribution will have the same frequency distribution as temperature emission for T, so that the emission per gram of the neutral element will be $\lambda k_\nu B(\nu, T)\,d\nu$, where λ is some constant.

Hence

$$\cos\theta\frac{dI_\nu(\theta)}{\rho\,dr}=-k_\nu I_\nu(\theta)+\lambda k_\nu B(\nu,T), \tag{6·11}$$

$$\frac{dH_\nu}{\rho\,dx}=-k_\nu J_\nu+\lambda k_\nu B(\nu,T), \tag{6·12}$$

and integrating (6·12) with respect to ν, then

$$\frac{dH}{\rho\,dx}=-\int k_\nu J_\nu\,d\nu+\lambda\int k_\nu B(\nu,T)\,d\nu. \tag{6·13}$$

But H is constant so that

$$\int_{\nu_0}^{\infty}k_\nu J_\nu\,d\nu=\lambda\int_{\nu_0}^{\infty}k_\nu B(\nu,T)\,d\nu, \tag{6·2}$$

the equation of radiative equilibrium in this case.

If the electron gas has no means of exchanging energy with the atmosphere except through transitions in this one continuous spectrum, one can equate the fraction of the energy absorbed which goes to electron velocities with the fraction of the capture spectrum which diminishes electron velocity energies, or

$$\int_{\nu_0}^{\infty}\frac{\nu-\nu_0}{\nu}k_\nu J_\nu\,d\nu=\lambda\int_{\nu_0}^{\infty}\frac{\nu-\nu_0}{\nu}k_\nu B(\nu,T)\,d\nu. \tag{6·3}$$

If now J_ν is $\omega S(\nu, T')$, where ω is some fraction and $S(\nu, T')$ is a temperature distribution corresponding to a (solar) temperature T', (6·2) and (6·3) can only be satisfied if $\lambda = \omega$ and $T = T'$. The re-emission is exactly the same as if the process were true scattering in a bound-bound transition: but the electrons, if they have no means of exchanging energy with the atmosphere except in this one band, must take up a velocity distribution appropriate to the colour temperature of the sun.

There will, of course, in any real atom be more than one continuous spectrum. In the case of O where ground state is $^3P_{0,1,2}$ there are excited states $^1D_2$ and $^1S_0$ with energies of 1·96 and 4·17 V. At sufficiently low dilution factors ω the forbidden transitions between these states and the ground state will become important, and the atom will transfer radiation from the ultra-violet band $O(^3P)\rightleftharpoons O^+(^4S)$ to the bands $O(^1S)\rightleftharpoons O^+(^4S)$ and $O(^1D)\rightleftharpoons O^+(^4S)$. The velocity distribution of the electrons will not be reduced unless high-velocity electrons show a greater tendency to fall with an excited state (rather than the ground state) than slow electrons.

If the temperature of the upper boundary of the earth's atmosphere falls much below 6000°, there must be some infra-red radiators, and if water vapour is dissociated at very great heights (see § 13) search must be made elsewhere for emission of low-temperature radiation. Two possibilities present themselves, which will be examined in succeeding sections. These are negative ions (§ 7) and dust (§ 8).

7. Although negative ions with their low dissociation potentials are very greatly dissociated into atoms and electrons in sunlight there will at all times be a small but finite negative-ion population. Since the continuous spectrum of the negative ion lies in the visible or infra-red, a very small negative-ion population may have a great effect on the temperature if there are no other low-frequency radiations available.

Let n_e, n^- and n be the numbers of electrons, negative ions and atoms per c.c. Then in dilute solar radiation the populations will be given by the modified Saha formula

$$\frac{n_e n}{n^-} = 2(2\pi m R)^{3/2} \frac{\varpi\omega}{8\pi h^3} e^{-\chi/RT'} \times F(T, T'), \tag{7·1}$$

where ϖ is the ratio of the statistical weights of the atom and negative ion, λ is the dissociation potential, T' the solar temperature, T the temperature of the atmosphere, and $F(T, T')$ is a factor which depends upon the form of the absorption coefficient of the negative ion as a function of frequency. In particular, if k_ν is proportional to ν^{-2}, $F(T, T') = T'T^{1/2}$. The subject is discussed very fully by Pannekoek (1930).

For given values of n_e and n we have n^- proportional to $e^{\chi/RT'}$. The infra-red emission per c.c. of such ions is proportional to

$$e^{\chi/RT'} \int_{\nu_0}^{\infty} k_\nu B(\nu, T)\, d\nu,$$

where $h\nu_0 = \chi'$ and the result of this integration will depend on the variation of k_ν with ν, but the most important term will be proportional to $e^{-\chi/RT'}$. Hence the emission per c.c. varies with

$$\exp\{\chi/RT' - \chi/RT\},$$

and if T' is considerably greater than T the emission is greater for small values of χ than for large ones. Thus the ion N^- whose dissociation potential is estimated at 0·04 V is more important than O^- whose potential is between 2 and 3 V.

As n_e decreases with increasing height, the ratio n^-/n decreases with increasing height, according to (7·1), so that in the limit where interstellar space is approached the influence of the negative ions disappears. However, at heights of 200 to 300 km. the negative ions will have a profound effect.

It is not possible to work out the radiation balance accurately because the proportion of nitrogen to oxygen, and the state of the nitrogen (whether molecular or atomic), is uncertain at great heights, and also because the absorption coefficient of negative nitrogen ions N^- and N_2^- are unknown. If N_2^- has a very low dissociation potential and an absorption coefficient nearly equal to that of N^-, it will make no difference to the calculation whether the nitrogen is in the molecular or atomic state, given the ratio of nitrogen particles to oxygen atoms. It will be supposed that negative ions showing a very low dissociation potential and exhibiting an absorption coefficient independent of frequency and equal to $4{\cdot}5\times10^{-18}$ cm.^{2} (the absorption coefficient of O) have to be dealt with. (It is known that the absorption coefficient of O^- is of order 10^{-18} cm.^{2}.)

Now if k_ν is independent of ν and if χ is small, the function $F(T, T')$ in (7·1) becomes $(T'T^{-\frac{1}{2}})^3$. Assuming a solar temperature of 6825° (for the visible spectrum) and an atmospheric temperature of 1000° for nitrogen it is found that

$$\frac{n_e n}{n^-} = 4{\cdot}7\times10^{15}. \tag{7·11}$$

Now assume that the atomic oxygen population is equal to the nitrogen population, and assume values of n and n_e. Then temperatures may be calculated from

$$n^- . acT^4 = n\int_{\nu_0}^{\infty} \omega h\nu^2(\nu-\nu_0)\, c^{-2} e^{-h\nu/RT'}\, d\nu. \tag{7·2}$$

Some results are shown in table 1. In computing this table T' has been taken as 6000° for the spectrum beyond 912 A. This is very uncertain, and the values of T are very sensitive to the adopted value of T'.

TABLE 1. TEMPERATURES BY EQUILIBRIUM BETWEEN NEGATIVE NITROGEN IONS AND OXYGEN ATOMS

$\log n$ assumed	10	8	6
$\log n_e$ assumed	5·7	4·7	3·7
$\log n^-$ from (7·11)	0·0	−3·0	−6·0
T, from (7·2)	1000°	1570°	3100°

Table 1 shows that the negative ions probably play a very important part in controlling the temperature. They are able, on the figures assumed, to radiate enough energy, compared with the intake of oxygen *atoms*, to pull the temperature down to 1000° or so. But as each oxygen molecule absorbs 4×10^5 times as much energy from full radiation at 6000° as an oxygen atom does, the negative ions would play no effective part in cooling a concentration of 10^{10} oxygen *molecules* per c.c. A discussion of the change-over from atomic to molecular oxygen with decreasing height in the atmosphere is given in § 14.

If it be supposed that the absorption coefficient of the negative ions is $4{\cdot}5 \times 10^{-18}$ and the molecular velocity temperature 1000°, and that there are 10^{10} relevant neutral particles and $10^{5{\cdot}7}$ electrons per c.c., each neutral particle will capture an electron once in about 18 days. Accordingly, since the neutral particles carry nearly all the kinetic energy, and can only lose it by capturing an electron and radiating a quantum, the atmosphere will not sensibly change its temperature from day to night. To obtain a day and night average, the temperatures shown in table 1 must be multiplied by $2^{-\frac{1}{4}}$, so that the value for $n = 10^{10}$ becomes 840°, etc.

It should be noticed that the absorption of ultra-violet energy is attributed to oxygen and not nitrogen. The nitrogen molecule does not appear to have a strong absorption continuum, and the nitrogen atom has an ionization potential of 14·40 V against 13·54 V for atomic oxygen, so that the nitrogen atom is the less efficient of the two in absorbing solar radiation.

8. Dust particles, if present in sufficient numbers, can cool the gases in the upper atmosphere. Suppose that there are N spherical dust particles per c.c. each of radius r and let $nr^2N = \sigma$. Then if there are n gas molecules per c.c. each of mass M the total number of molecules which strike dust particles per c.c./sec. is

$$2\sigma n\left(\frac{2RT}{\pi M}\right)^{\frac{1}{2}},$$

the average kinetic energy being $2RT$ (swifter particles striking more often than slower ones). As usual T is the parameter which characterizes the Maxwellian velocity distribution of the gas. If the dust particles are maintained by their radiation balance at some temperature T', the gas molecules will rebound from the dust with an average kinetic energy $2RT''$, where

$$T - T'' = \alpha(T - T'),$$

and α is called the accommodation coefficient. Hence the net interchange of energy per c.c./sec. from the gas to the dust is

$$2\sigma n\left(\frac{2RT}{\pi M}\right)^{\frac{1}{2}} 2\alpha R(T - T'),$$

and if E is the energy absorbed per c.c./sec. by each molecule from the radiation field, and if the dust is the principal cooler, then

$$E = 4\alpha\sigma R\left(\frac{2RT}{\pi M}\right)^{\frac{1}{2}}(T - T'). \qquad (8{\cdot}1)$$

The value of σ may be estimated from the brightness of the zodiacal light. According to Elvey & Roach (1937) the brightness of one square degree of the zodiacal light at a point on the ecliptic 90° from the sun is equal to 200 stars of the tenth magnitude. Let the albedo of the dust be ϵ. Then

$$\epsilon\sigma = 5 \times 10^{-17} \times \phi,$$

where ϕ is a factor depending on the law of distribution of σ as a function of the distance from the sun, having the value of 0·67 if σ is inversely proportional to the distance of 1·25 if it is proportional to the inverse fourth power, so that one may assume $\phi = 1$. The value of the albedo may be estimated from that of the moon (0·07) and of Ceres (0·06) rather than from the albedos of planets which have atmospheres. Taking $\epsilon = 0{\cdot}07$ and $\phi = 1$, then $\sigma = 7 \times 10^{-16}$. Now it is easily seen that in the upper atmosphere the intake of solar energy by oxygen atoms is unable to heat the dust materially so that $T' = 233°$ (by day). If it be supposed that oxygen atoms constitute the fraction f of the total population which conveys energy from gas to dust and differences of mass in the gas particles be ignored, then

$$\frac{Ef}{\alpha\sigma} = 4R^{\frac{3}{2}}\left(\frac{2}{\pi M}\right)^{\frac{1}{2}} T^{\frac{1}{2}}(T-233), \qquad (8{\cdot}2)$$

and E can be calculated from the absorption coefficient for oxygen.

The resultant values of T are highly dependent on the assumed temperature of the sun in the ultra-violet, and on the values of f, α and σ. With $f = \frac{1}{2}$, $\alpha = \frac{1}{2}$, $\sigma = 7 \times 10^{-16}$ and a solar temperature of 6000° we get $T = 4000°$. With these figures the dust is less efficient as a cooler than negative nitrogen ions at a level in the atmosphere when $n = 10^6$ atoms per c.c., according to the preceding section.

It should, however, be emphasized that both calculations may be substantially revised when relevant coefficients and conditions become better known. In default of other infra-red radiators, however, one or other of these types of particle, or possibly both acting at different heights, must determine the molecular velocity temperature at great heights in the earth's atmosphere.

9. It is not possible to proceed further without examining the dissociation of the molecules in the atmosphere into their constituent atoms as a function of height. In full equilibrium at a temperature T the dissociation of a homonuclear diatomic molecule is given by

$$\left(\frac{n_1^2}{n_2}\right)_T = \frac{G}{2hr_0^2}\left(\frac{MRT}{\pi}\right)^{\frac{1}{2}} e^{-\chi/RT}, \qquad (9{\cdot}1)$$

where n_1, n_2 are the numbers of atoms and molecules per c.c., M the mass of each atom and G, r_0 and χ are constants, the latter being the dissociation potential. This expression is derived, for example, in Rosseland's *Theoretical Astrophysics*, where values of the constants for various molecules are given. For oxygen, $G = 27$, $r_0 = 1{\cdot}20 \times 10^{-8}$ cm., and $\chi = 5{\cdot}09$ V. Then from (9·1) with $T = 1000°$, $n_1^2/n_2 = 0{\cdot}319$ and for $T = 300°$ the value is 3×10^{-61}. Just as Saha's equation leads to Milne's well-known relation (1924*a*) between the continuous absorption coefficient of an atom and the probability that an ion will capture an electron, so equation (9·1) (which is called the isochore) leads to a relation between the absorption coefficient of the molecule and the probability of recombination between two atoms, which is mentioned here because the oxygen absorption coefficient already quoted in earlier sections leads to a mean life of the oxygen atoms of the order of 25 days with $n_1 = 10^{12}$, showing that dissociation may safely be averaged over day and night.

The effect of departures from thermodynamical equilibrium on ionization (which is very similar in this respect to dissociation) was first examined by Woltjer and applied to stellar atmospheres by Pannekoek (1926, 1930). The same ideas have been applied specifically to dissociation by Majumdar (1938). The form of the result will depend on the variation of k_ν with ν. If one supposes, for example, that k_ν is proportional to $\nu^{-2}(\nu-\nu_0)$ (where $h\nu_0 = \chi$) for all $\nu > \nu_0$ and zero for $\nu < \nu_0$, then (9·1) becomes

$$\frac{n_1^2}{n_2} = \frac{G\omega}{16\pi^{\frac{3}{2}}} \frac{(MRT)^{\frac{1}{2}}}{hr_0^2} \left(\frac{T'}{T}\right)^2 e^{-\chi/RT'}, \tag{9·2}$$

where T' is as before the solar temperature. With $T' = 6825°$ and $\omega = 5{\cdot}41 \times 10^{-6}$, then for oxygen, with $T = 1000°$, $n_1^2/n_2 = 2{\cdot}73 \times 10^{16}$ and with $T = 300°$, $n_1^2/n_2 = 1{\cdot}66 \times 10^{17}$, in great contrast to the values given by (9·1). Except at the top of the atmosphere, the radiation is not simply dilute temperature radiation, but is weakened by absorption by day and is absent at night. Since the life of the dissociated atom is large compared with 24 hr., the dissociation may be calculated by taking an average over day and night, and as regards absorption, the value of n_1^2/n_2 must be multiplied by an appropriate factor before this average is taken.

Ozone will be ignored and attention confined to heights above the ozone layer. Besides radiative recombination and dissociation consideration will be given to the effects of three-body collisions of type

$$O_2 + X \rightleftharpoons O + O + X.$$

If n_x is the number of X particles per c.c. the number of dissociations per c.c./sec. due to collisions is proportional to $n_2 n_x$, and the number of recombinations is proportional to $n_1^2 n_x$. Then if the day and night variation of dissociation is small secular equilibrium follows, or

$$n_2(q + \Sigma b_x n_x) = n_1^2(a + \Sigma \lambda_x n_x),$$

the summation being over all possible types of third body X. q is the number of dissociating quanta absorbed per second by the molecule. In thermodynamic equilibrium the collision processes balance, or

$$(n_2 \Sigma b_x n_x)_T = (n_1^2 \Sigma \lambda_x n_x)_T,$$

so that if dilute radiation greatly increases the value of n_1^2/n_2 over $(n_1^2/n_2)_T$, then $n_1^2 \Sigma \lambda_x n_x$ is much greater than $n_2 \Sigma b_x n_x$. Under such circumstances the dissociation equilibrium equation is

$$n_2 q = n_1^2(a + \Sigma \lambda_x n_x). \tag{9·3}$$

This may be compared with a formula used by Wulf & Deming (1936) in calculating the dissociation of oxygen in the atmosphere. Their equation (8) in the present notation is

$$n_2 q = n_1^2 \Sigma \lambda_x n_x, \tag{9·4}$$

which is clearly applicable in regions where the recombination through collisions is more important than radiative recombination. It is clear, however, that (9·4) will give too small a molecular concentration when n_x becomes small—that is, at great heights.

10. Now consider the gravitational equilibrium of an isothermal dissociating atmosphere in which there is no mixing by vertical transport. Choosing the axis of x vertical and positive upwards, the equation of hydrostatic equilibrium is

$$dp = -g\rho\, dx, \tag{10·1}$$

where p is the total pressure and ρ the total density. Consider an atmosphere consisting only of O and O_2 reacting with one another. Let M be the mass of the oxygen atom and let there be n_1 oxygen atoms and n_2 oxygen molecules per c.c. Then

$$p = (n_1+n_2)\, RT, \quad \rho = M(n_1+2n_2).$$

If there is no absorption of the dissociating radiation and if collisions may be neglected, then

$$qn_2 = an_1^2, \tag{10·2}$$

where q and a are constant. These conditions apply to the top of the atmosphere. Equation (10·1) becomes

$$(dn_1+dn_2)\, RT = -Mg\, dx(n_1+2n_2).$$

But from (10·2) $q\, dn_2 = 2an_1 dn_1$. Accordingly

$$dn_1(q+2an_1)\, RT = -Mg\, dx(q+2an_1),$$

$$\frac{dn_1}{n_1} = -\frac{Mg}{RT}\, dx,$$

and writing $RT/Mg = H$, then

$$n_1 = Ae^{-x/H}, \quad n_2 = Be^{-2x/H}. \tag{10·3}$$

The constants A and B are related by $qB = aA^2$. Each constituent is distributed vertically with height exactly as if the reaction did not take place. This result applies to any number of reacting constituents. It may be proved in the case of full equilibrium from a general theorem due to Willard Gibbs (1875), and the result was applied to gravitational equilibrium in a reaction of type (10·2) with q/a constant by Milne (1924b). The result, of course, does not hold if n_1 and n_2 are connected by the more general relation

$$n_2(q+\Sigma b_x n_x) = n_1^2(a+\Sigma \lambda_x n_x),$$

unless (1) there is full equilibrium or (2) collisions are much less important than radiation (equation (10·2) above) or (3) collisions are much more important than radiation. In the earth's atmosphere case (2) applies to great heights and case (3) to levels near the ground. The transition is examined in the next section.

11. The term q is weakened by absorption as one goes downwards in the atmosphere, but to obtain a solution it will be supposed that q is constant until the transition from atomic to molecular oxygen has been passed. With q constant a simple solution of (10·1) and (10·2) is obtained if the third bodies X are the oxygen molecules and atoms themselves, and if they are equally efficient in three-body collisions, for then (10·2) becomes (9·3), in this case

$$qn_2 = n_1^2(a+\lambda(n_1+n_2)). \tag{11·1}$$

The solution of (10·1) and (11·1) is

$$n_1(1-\lambda q^{-1}n_1^2)^{-\frac{1}{2}} = Ce^{-x/H}, \quad qn_2 = (a+\lambda n_1)\,C^2e^{-2x/H}. \tag{11·2}$$

For n_1 small, then $n_1 = Ce^{-x/H}$ and $qn_2 = aC^2e^{-2x/H}$, as in (10·3). As x decreases, n_1 increases until it approaches the value $q^{\frac{1}{2}}\lambda^{-\frac{1}{2}}$ asymptotically. At low levels n_2 is approximately equal to

$$\left\{\frac{a}{q}+\left(\frac{\lambda}{q}\right)^{\frac{1}{2}}\right\}C^2e^{-2x/H}.$$

Now q/a is equal to the value of n_1^2/n_2 in (9·2), and q is known separately from the absorption coefficient. Further, $\left\{\frac{a}{q}+\left(\frac{\lambda}{q}\right)^{\frac{1}{2}}\right\}C^2$ is the number of oxygen molecules per c.c. at ground level. Hence the distribution of oxygen molecules and atoms in the atmosphere is known if λ is known. The number of oxygen atoms per c.c. at ground level is not given correctly by these equations, as absorption will weaken q and finally reduce it to the level where the collision term $n_2\Sigma b_x n_x$ (omitted from (11·1)) is important. Thus the ground value of n_1 is given by (9·1) with the ground temperature of the atmosphere. If the absorption of the ultra-violet radiation only becomes important below the level where collisions have reduced the oxygen to the molecular state, the values of n_2 will be correct throughout the atmosphere and those of n_1 correct in and above the transition layer. It is, however, wrong to suppose that atoms are as efficient as molecules in absorbing energy in three-body collisions. The molecule with its vibrational and rotational degrees of freedom offers far greater possibilities of absorbing energy through resonance. Two cases arise: (*a*) where the molecules of the substance itself are the most important contributors to $\Sigma\lambda_x n_x$ or $\Sigma\lambda_x n_x = \lambda n_2$, and (*b*) where the molecules of another gas are the most important contributors, or

$$\Sigma\lambda_x n_x = \lambda n' e^{-2x/H}.$$

In case (*a*), which presumably applies to nitrogen, the equation (9·3) becomes

$$qn_2 = n_1^2(a+\lambda n_2),$$

which, together with (10·1), has the solution

$$\left.\begin{aligned} n_1\left\{\frac{1+2\alpha n_1^2-\mu n_1^2}{1-\mu n_1^2}\right\}\left\{\frac{\sqrt{(\alpha^2+\mu)}+\alpha-\mu n_1}{\sqrt{(\alpha^2+\mu)}-\alpha+\mu n_1}\right\}^{\frac{\alpha}{2\sqrt{(\alpha^2+\mu)}}} &= De^{-x/H}, \\ n_2 &= n_1^2(\alpha+\mu n_2), \end{aligned}\right\} \tag{11·3}$$

where $a/q = \alpha$, $\mu = \alpha\lambda$ and D is a constant of integration.

Case (*b*) applies to oxygen with nitrogen molecules as the third bodies. Strictly speaking the value of H is not the same for the two molecules, but recognition of this produces analytical complications in expressions which are already cumbrous. Ignoring this point (9·3) becomes

$$qn_2 = n_1^2(a+be^{-2x/H}). \tag{11·4}$$

It is convenient to write $2an_1/q = y$. Then the solution of (11·4) and (10·1) is

$$ay^2 = 2\{E-\ln(1+y)+y-\tfrac{1}{2}y^2\}\,be^{-2x/H}, \tag{11·5}$$

n_2 being derived from y by (11·4), and E being the constant of integration. It is not obvious that (11·4) and (11·5) are essentially similar to (11·2), but in (11·5) as one goes downwards y approaches the value $y = y_1$, where

$$\ln(1+y_1)-y_1+\tfrac{1}{2}y_1^2 = E.$$

At the top of the atmosphere where y is small

$$y = (2\lambda E)^{\frac{1}{2}}\,e^{-x/H},$$

and at the bottom of the atmosphere

$$n_2 = \frac{qby_1^2}{4a^2}e^{-2x/H}.$$

Although the actual atmosphere is not isothermal, calculations for an isothermal atmosphere at $T = 300°$ probably represent the first 150 km. of the atmosphere well. From Wulf & Deming's estimate (1938) of the collision factor b may be taken in (11·4) at 10^{-12}. If the solar temperature applicable to 2000 A be estimated at 6300°, then $q = 6\times10^{-5}$ quanta per sec. per molecule for oxygen. With $T' = 6300$ and $T = 300$ then $q/a = n_1^2/n_2 = 4{\cdot}9\times10^{16}$ from (9·2). Hence $a = 1{\cdot}26\times10^{-21}$. Values of $\log n_1$ and $\log n_2$ computed from (11·4) and (11·5) are shown in table 2. These values are plotted in figure 1 and joined by full lines. The pecked lines show the distribution of O and O_2 at great heights if the temperature is 1000° at those heights and if the value of n_1 for O is 2×10^{10} (§ 14) and the dotted line shows values obtained in § 12 (table 3). In § 5 it was seen that the Schumann-Runge absorption is reduced to one-half below 2×10^{17} molecules. With the conditions assumed, this will occur rather above 100 km., and the assumption that q can be taken constant requires modification. This would result in greater molecular and smaller atomic populations than those shown in table 2. This effect will be increased by temperatures higher than 300°.

It is less easy to estimate the dissociation of nitrogen as the spectrum of N_2 is not so well known as the spectrum of O_2. On account of its greater dissociation potential (9 V) N_2 is much less dissociated than O_2.

TABLE 2. DISSOCIATION OF OXYGEN IN AN ISOTHERMAL ATMOSPHERE

H, km.	83	98	110	122	133	170	206	242	278
$\log n_1$	13·23	13·16	12·99	12·69	12·39	11·39	10·39	9·39	8·39
$\log n_2$	14·10	13·15	12·17	10·90	9·69	6·19	4·09	2·09	0·09

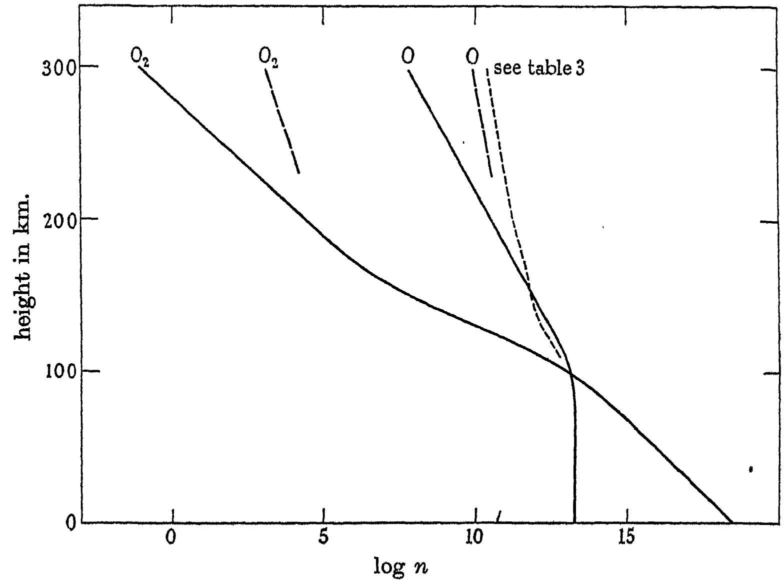


FIGURE 1. Distribution of O and O_2 with height. —— Isothermal atmosphere at 300° C (table 2). ---- temperature increasing from 300 to 1000° C (table 3). — — — values based on collision frequencies (§ 14).

12. As has been seen in earlier sections, above 100 km. or so (the limit to which the Schumann-Runge radiation reaches) the temperature very probably increases, in a manner depending upon the distribution of the sources of ultra-violet absorption and infra-red emission. To examine the effect of the temperature gradient on the distribution of population with height, assume as a first approximation that the atmosphere consists of only one constituent and that the temperature gradient is proportional to the population at any level. Let there be n particles per c.c., each of mass M. The $\rho = nM$ and $p = nRT$. Now it has been supposed that

$$RdT = \sigma g\rho dx, \qquad (12{\cdot}1)$$

where σ is some constant. In addition

$$dp = -g\rho dx. \qquad (12{\cdot}2)$$

These conditions are supposed to hold from $x = x_1$ to $x = \infty$. Let $T = T_1$ and $n = n_1$ at $x = x_1$, and let $RT_1/Mg = H_1$. Then the solution of (12·1) and (12·2) is

$$\left.\begin{aligned} &(\sigma n+1)\,T = (\sigma n_1+1)\,T_1, \\ &\frac{x-x_1}{H}+\left(\frac{T}{T_1}-1\right)+(\sigma n_1+1)\ln\frac{nT}{n_1T_1} = 0. \end{aligned}\right\} \qquad (12{\cdot}3)$$

Suppose, for example, that $n_1 = 10^{13}$ and $T_1 = 300°$ at $x_1 = 100$ km., and that $H_1 = 15$ km. Let us choose the value of σ so as to make T tend to the limiting value 1000°. Then the temperature and population will be distributed with height as is shown in table 3.

TABLE 3. TEMPERATURE INCREASING UNIFORMLY WITH NUMBER OF PARTICLES ABOVE 100 KM.

$\log n$	13·0	12·48	12·0	11·48	11·0	10·48
H, km.	100	112	140	187	222	282
T	300	589	810	935	977	993

13. On account of its importance as a source of infra-red radiation, water vapour requires special discussion. The equilibrium between N water-vapour molecules per c.c. on the one hand and n hydrogen atoms and n' oxygen atoms per c.c. on the other is given by a formula similar to (9·1), namely,

$$N = n^2 n' K(T),$$

and by the Willard Gibbs-Milne argument all the constituents are distributed vertically in an isothermal atmosphere in gravitational equilibrium just as if no dissociation took place or

$$N = Ce^{-x/H}, \qquad (13{\cdot}1)$$

the value of H for water vapour being nearly equal to that of atomic (not molecular) oxygen. In full equilibrium, therefore, the water-vapour population would *gain* over the molecular oxygen population with increasing height. This argument, however, overlooks the escape of hydrogen from the earth's atmosphere. The view was put forward by Chapman & Milne (1920) that there is no appreciable amount of free hydrogen in the atmosphere. This is generally accepted, and, if true, must have as a consequence the result that water vapour falls off with increasing height far more rapidly than is given by equation (13·1). If the temperatures at great heights were known one could calculate the rate at which hydrogen atoms escape from the atmosphere from a given population distribution; and at the same time work out the populations of hydrogen atoms and water-vapour molecules at all heights if a single constant were known—the rate at which water-vapour molecules are transported over unit area of the tropopause. This constant is related to the total number of water-vapour molecules in a column of unit cross-section above the tropopause. As has been suggested earlier, it appears to be within the reach of observation in the near future. The importance of this constant to ionospheric theory can hardly be overestimated.

14. The physical arguments in §§ 6 to 13 are, from the point of view of the earlier sections, arguments about the absorption coefficient in the infra-red and ultraviolet as a function of height. The problem is complicated by the fact that one requires a knowledge of the temperature to determine the constituents of the atmosphere, so that even with the fullest knowledge of the atomic coefficients involved one would be compelled to proceed by successive approximation from an assumed temperature distribution. Since many quantities cannot at present be evaluated accurately, this will not be done. One may, however, inquire whether there are two substantially distinct regions heated by molecular oxygen and atomic oxygen and separated by a temperature minimum.

Possibly the soundest method of estimating the atomic population at great heights is an appeal to the observed collision frequency. According to Farmer & Ratcliffe (1935), an electron makes $1\cdot3 \times 10^3$ collisions per second in the F_2 region. This experimental result is confirmed by Martyn & Pulley (1936). By comparison with the collision frequency in the E region the latter authors deduce that the temperature in the F region is of order 1000°. We may for definiteness suppose that at a height of 250 km. the molecular population is 4×10^{10}, and that one-half of this is atomic oxygen (the result being either atomic or molecular nitrogen). Assuming that at this height one is well above the levels where Schumann-Runge radiation is substantially absorbed, then from (9·2) and making an allowance for day and night averaging, $n_2 = 5 \times 10^4$. Now at a temperature of 1000° the value of H for oxygen is 52·2 km., so that if the atmosphere is isothermal above 250 km. we have by (10·3) $1\cdot05 \times 10^{17}$ oxygen atoms and $3\cdot10^{11}$ oxygen molecules in a column of unit cross-section above this level. Since for O, $k = 4\cdot5 \times 10^{-18}$, the atomic oxygen radiation is just becoming appreciably absorbed at this point, being reduced to $e^{-0\cdot45}$ or 0·64 of its incident value.* With a solar-energy curve of 6000° each oxygen molecule absorbs 4×10^5 times as much energy as an oxygen atom, and the two intakes of energy are roughly equal at 250 km. with the figures which we have adopted.

Going upwards the temperature must increase, as the electron density decreases (the layer is the maximum electron density), so that atomic population gains over negative-ion concentrations with increasing height. Going downwards the electron population again decreases, so that the oxygen molecular population will gain over the negative-ion population. Hence unless water vapour is present in sufficient quantity to compete with the infra-red radiation from negative ions at the 250 km. level, that level is, on the figures adopted, a minimum of the temperature. Going downwards from 250 km. the temperature increases until water vapour is encountered. The fact that the recombination coefficient for the F_1 region (say 180 km.) is much greater than that for the F_2 region (say 250 km.) strongly suggests that there are many more negative (oxygen) ions in the lower region, and therefore that the temperature in the F_1 region is considerably lower than the temperature in the F_2 region.* This may well be due to an appreciable water-vapour population in the lower region.

---

* If the F_2 region is formed by the ionization of O to O^+ one would expect this radiation to be reduced to e^{-1} of its initial value at the level of maximum ionization.

The author acknowledges with pleasure the benefit he has derived from discussing ionospheric problems with Dr D. F. Martyn and Dr C. W. Allen.

REFERENCES

Angenheister 1932 *Terr. Magn. Atmos. Elect.* **37**, 431.
Appleton & Naismith 1935 *Proc. Roy. Soc.* A, **150**, 685.
Chapman & Milne 1920 *Quart. J. R. Met. Soc.* **46**, 357.
Elvey & Roach 1937 *Astrophys. J.* **85**, 213.
Farmer & Ratcliffe 1935 *Proc. Roy. Soc.* A, **151**, 370.
Godfrey & Price 1937 *Proc. Roy. Soc.* A, **163**, 228.
Gold 1909 *Proc. Roy. Soc.* A, **82**, 43.
Gowan 1928 *Proc. Roy. Soc.* A, **120**, 655.
Gowan 1930 *Proc. Roy. Soc.* A, **128**, 531.
Ladenburg & van Voorhis 1933 *Phys. Rev.* **43**, 315.
Lindemann & Dobson 1922 *Proc. Roy. Soc.* A, **102**, 411.
Majumdar 1938 *Indian J. Phys.* **21**, 75.
Martyn & Pulley 1936 *Proc. Roy. Soc.* A, **154**, 455, 463.
Milne 1922 *Phil. Mag.* **44**, 872.
Milne 1924*a* *Phil. Mag.* **47**, 209.
Milne 1924*b* *Proc. Camb. Phil. Soc.* **22**, 493.
Pannekoek 1926 *Bull. Astr. Insts Netherlds*, **3**, 207.
Pannekoek 1930 *Handbuch der Astrophysik*, **3**, 288.
Willard Gibbs 1875 *Sci. Papers*, **1**, 144, 171.
Wulf & Deming 1936 *Terr. Magn. Atmos. Elect.* **41**, 302.
Wulf & Deming 1938 *Terr. Magn. Atmos. Elect.* **43**, 283.

---

* I owe this remark to Dr Martyn.

Atmospheric tides in the ionosphere

I. Solar tides in the F_2 region

By D. F. Martyn, D.Sc.,
Australian Council for Scientific and Industrial Research

(*Communicated by Sir Edward Appleton, F.R.S.—Received* 23 *April* 1946)

A theory, based on solar tides, is advanced to explain the anomalous seasonal, diurnal and geographical variations of F_2 region ionization. It is shown that the horizontal winds due to these tides must cause electrons to move along the lines of the earth's magnetic field. The resultant motion has a vertical component. Account is taken of polarization of the medium by the 'dynamo' electric forces. Owing to viscosity the vertical motion decreases upwards in the F_2 region. Application of the equation of continuity shows that the F_2 region becomes greatly distorted. A 'longitude effect' is found to arise by reason of the asymmetry of the earth's magnetic field. The theory is used to explain the high F_2 ionization densities found in low latitudes, and the high values of h'_{F_2} at noon near the equator. It is also used to explain the afternoon and night-time increases in ionization found at certain locations. It is suggested that the effective recombination coefficient in F_2 is much lower than the generally accepted values. It is shown that Appleton & Weekes's evidence of lunar tidal effects in the E region does not conflict with the 'dynamo' theory of magnetic variations or with Pekeris's calculations.

Observational evidence of the existence of solar tides in the F_2 region is presented.

1. Introduction

Soon after the experimental discovery of the ionized regions of the upper atmosphere Chapman (1931) proposed a theory of their production. According to this theory, if the ionization ultimately disappears as if by a recombination law, then a given radiation acting on a given atmospheric constituent will produce a maximum density of ionization N(max.) at a certain level in the atmosphere. This level will depend on the absorption coefficient of the constituent for the radiation, on the temperature of the atmosphere, and on the solar zenith distance χ. When N(max.) reaches a temporal maximum, which usually happens very soon after noon, then theoretically it should attain a value proportional to $\cos^{\frac{1}{2}}\chi$.

This theory has been applied by Appleton to a number of ionospheric problems, and with particular success to the E and F_1 regions. All attempts to apply it to the uppermost and most densely ionized region, the F_2 region, have met with serious difficulties. For example, the noon value of N(max.) in this region is usually much greater in winter than in summer, while the distribution of noon N(max.) over the globe is much more complicated than that indicated by the $\cos^{\frac{1}{2}}\chi$ law. (In a recent unpublished note on the 'Morphology of the F_2 region', Appleton has produced evidence that noon N(max.) depends to some extent on magnetic latitude.)

A suggested explanation of the low values of summer N(max.) put forward independently by Hulbert (1934) and Appleton (1935) is that the F_2 region is much hotter (some ten times) in summer than in winter. Martyn & Pulley (1936), however, have advanced evidence that the F_2 region attains a temperature of the order 1000° C all the year round, a conclusion which receives strong theoretical support from considerations of radiative equilibrium in that region (Godfrey & Price 1937; Woolley 1946).

It appears desirable, therefore, to explore other possible explanations of the peculiarities of the F_2 region. Up till now, with the exception of the early work of Hulbert (1929), theoretical treatments of the ionosphere have assumed that the upper atmosphere was essentially static. We explore here certain aspects of a dynamical theory of the ionosphere. In this series of papers the view is propounded that the F_2 region is subject normally to marked distortion by the vertical transport of ionization with velocities varying both with height and time. In this way the observable quantities N(max.), $h^{\text{max.}}$, the height at which N(max.) occurs, and h', the minimum equivalent height, can be considerably affected, even though the total ionization in the region remains unaltered. (In general, however, the total ionization in the region *will* be affected if the effective recombination coefficient α' is a function of height.)

Evidence will be adduced that such vertical movements of ionization are caused by the air movements associated with the solar atmospheric tide. Although this tidal motion is predominantly horizontal, ions in the upper atmosphere are constrained to move along the lines of the earth's magnetic force, which has a vertical component except at the magnetic equator. Looked at in another way, which it will be shown later leads to an identical result, the horizontal movement of ionized air across the *vertical* component of the earth's magnetic field gives rise to horizontal electric forces, as in the 'dynamo' theory of magnetic variations originally due to Balfour Stewart (1882) and developed by Schuster (1908) and Chapman (1919). Now these horizontal electric forces, acting across the *horizontal* component of the earth's field, will give rise to a vertical transport of ions, all kinds of which will move in the same direction. It is a further feature of the 'dynamo' theory that a certain distribution of electrostatic charge must occur over the earth's atmosphere such as to produce a horizontal electric field which in low latitudes opposes the weaker dynamo field, so closing the current circuit in each hemisphere. It follows that even near the magnetic equator there exist electrodynamical forces capable of causing vertical transport of ionization, though in these regions such movement will have a sense opposite to that happening simultaneously at high latitudes.

In the present paper data from a number of ionospheric observatories are analysed, and a number of phenomena presented which show the existence of a large *semi-diurnal* harmonic in N(max.), $h^{\text{max.}}$ and h'. An electrodynamical theory of the vertical transport of ions will then be developed, starting from the general surface harmonic term ψ_n^σ in the velocity potential of the solar tide. Although modern tidal theory casts doubt on the existence of a true velocity potential, the main features of the actual tide could be represented by a small number of such terms. It will be found that the asymmetry of the earth's magnetic field with respect to the geographic poles introduces longitude terms whose effects are specially important between latitudes 25 and 55°. An examination of the effect on N and $h^{\text{max.}}$ of the vertical gradient of velocity will then be made. In conclusion, the application of the theory to the interpretation of the main anomalies in F_2 region behaviour will be discussed.

In subsequent papers in this series it is intended to apply the theory in detail to the elucidation of lunar tidal effects, and to the mechanism of formation of the so-called abnormal or sporadic E ionization.

2. EVIDENCE SUGGESTING THE EXISTENCE OF TIDAL CIRCULATIONS IN THE F_2 REGION

2·1. Before proceeding to develop this theory quantitatively in detail, it is desirable to set out certain data which appear to suggest the importance of tidal circulations in the ionosphere.

In a recent communication Martyn (1945) has presented evidence that noon h'_{F_2} at Huancayo, near the magnetic equator, shows a seasonal variation which is like that recorded at northern hemisphere observatories for the months when the sun is north of Huancayo, and like that at southern observatories when the sun is to the south. The somewhat more complicated seasonal variation of noon N(max.) shows a similar phenomenon. This behaviour, which is similar to that of the diurnal variation of magnetic declination at Huancayo (Fleming 1939), suggests that the variations of h'_{F_2}, like the magnetic variations, may be due to hemisphere-wide tidal circulations.

2·2. Some further evidence suggestive of tidal action in the F_2 region is derived from a study of the times at which h' reaches maximum values in the course of its diurnal variation. It is a striking feature of the results from many ionospheric observatories that h' and $h^{\text{max.}}$ attain maximum values not only near *noon* but also near *midnight*. This is clearly demonstrated in figure 1, which shows the diurnal variations of h' and $h^{\text{max.}}$ for Mount Stromlo (Canberra), averaged over the 4 years 1941–4. While various hypotheses, such as that of heating, might be invoked to account for a noon maximum, it is difficult to see why another maximum should ensue near midnight. On the other hand, if this tidal hypothesis is correct, a 12 hr. periodicity in such phenomena should be expected.

To examine this point more closely the monthly mean hourly values if h'_{F_2} for Mount Stromlo have been selected, which have the advantage of being available without interruption for 100 months, from April 1937 to date. The mean diurnal curve for each month was plotted and examined for maxima. In general, one maximum only was found during the night hours, and the time of its occurrence was noted to the nearest half-hour. During the day-time there was normally only one maximum present, and its time also was noted. Occasionally, particularly in the summer months, the curve took a small dip near the normal maximum, giving two maxima close together. In this case the time recorded was the mean time of the two close maxima. This method of treating the data, though obviously crude, suffices to demonstrate the point. In figure 2 are plotted the times of the day and night maxima respectively. It will be seen, first, that over the whole period the average day-time maximum occurs very nearly at noon, while the average night maximum occurs very close to midnight. There is thus an average periodicity of almost exactly 12 hr. in the occurrence of the maximum over 100 months. Secondly, it will be observed that the time of occurrence of each maximum shows a seasonal swing, being about an hour before midday (or midnight) in summer, and an hour after midday (or midnight) in winter. The fact that the curves swing in phase from summer to winter, constantly preserving a time difference of 12 hr., gives further support to the hypothesis that one is concerned with a semi-diurnal solar tide.

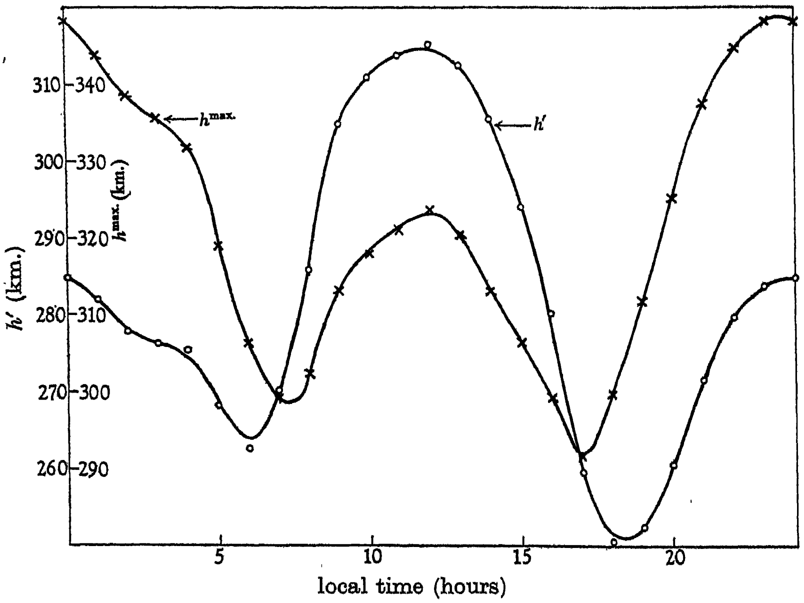


FIGURE 1. Average diurnal h'_{F_2} and $h^{\max.}_{F_2}$, Canberra, 1941–4 inclusive.

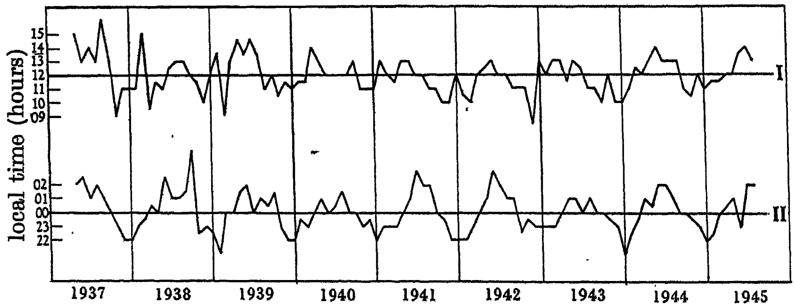


FIGURE 2. Monthly average time of occurrence of maximum h'_{F_2} at Canberra, 1937–45. Curve I, day-time maximum; curve II, night-time maximum.

A more accurate study of this point is provided by harmonic analysis of the diurnal variations. For this purpose $h^{\text{max.}}_{F_2}$, as determined by the method of Booker & Seaton (1939), has been selected. This quantity is apparently the most definite index of height variations in the F_2 region available without excessive labour. The data subjected to analysis are again for Mount Stromlo, consisting of the monthly mean hourly values of $h^{\text{max.}}$ average over the years 1941–4 inclusive. In table 1 the results of this analysis are presented for the first four harmonics, whose amplitudes are represented by P_1, P_2, etc. km., and times of maximum by t_1, t_2, etc. hr. after local midnight.

It will be observed from the table that P_2 is by far the most consistent harmonic, in both amplitude and phase. P_1, although larger than P_2 in winter, is subject to such large changes of amplitude and phase as would make one hesitate to suggest the presence of a *diurnal* tidal influence. In this connexion it will be realized that if the semi-diurnal tide affects $h^{\text{max.}}$ differently by day than by night (a conclusion which will be reached subsequently by theoretical reasoning), then it would give rise to a term in P_1. It will also be noticed that the phase of P_2 swings from winter to summer in a manner similar to that exhibited by h' in cruder if more graphic form in figure 2.

Table 1. Harmonic analysis of average diurnal variation $h^{\text{max.}}_{F_2}$ at Mount Stromlo for years 1941–4 inclusive

	amplitude in km.				time of max. in hr. after midnight			
month	P_1	P_2	P_3	P_4	t_1	t_2	t_3	t_4
Jan.	2·07	20·4	5·8	6·7	7·9	11·6	5·2	3·3
Feb.	5·40	19·8	7·6	6·2	22·6	11·7	3·9	4·1
Mar.	23·2	16·5	5·1	6·4	0·2	12·2	4·8	4·2
Apr.	30·4	14·4	6·6	2·9	0·5	12·4	4·2	1·5
May	31·5	14·4	4·8	3·8	0·2	12·8	3·3	1·5
June	31·1	13·6	3·9	1·7	0·3	12·7	4·0	0·6
July	26·3	13·4	7·1	3·4	0·2	12·5	3·7	0·6
Aug.	22·7	14·0	6·9	3·1	1·0	11·8	3·9	5·9
Sept.	15·8	14·7	7·7	2·6	1·0	10·8	4·1	4·5
Oct.	10·6	16·1	4·3	2·8	3·6	10·9	5·8	3·8
Nov.	5·75	18·4	1·1	6·6	3·3	10·9	6·4	3·2
Dec.	4·13	14·9	3·2	9·2	22·1	11·7	5·6	3·2

2·3. The evidence so far advanced relates only to h'_{F_2}, though mention has been made in § 2·1 of the fact that N(max.) at Huancayo shows a behaviour similar to that for h'. It is common knowledge that there is a close relation between h' and N(max.) for the F_2 region, large values of h' being usually accompanied by low values of N(max.). It must be expected, however, that any semi-diurnal *tidal* term in N(max.) will be overshadowed by the large diurnal *ionizing* term due to direct solar radiation. In seeking the former term it is clearly preferable to compare the diurnal variation of N(max.) at two or more latitudes so chosen that while the ionizing term (proportional to $\cos^{\frac{1}{2}}\chi$) varies little, the tidal term (whose variation with latitude is developed in § 4·2 below) varies markedly. Three suitable sites for this purpose are Cape York, Queensland (lat. 11·0° S), Brisbane, Queensland (lat. 27·5° S) and Canberra, A.C.T. (lat. 35·3° S). The monthly mean hourly values of $f^0_{F_2}$ for the equinoctial month March 1945 are selected. This month is chosen for the reason that, the sun being over the equator, asymmetrical terms are absent in the tidal potential. In figure 3 are plotted two curves: (1) the difference between corresponding hourly values of $f^0_{F_2}$ at Cape York and Canberra, and (2) the difference between the same quantities for Brisbane and Canberra.

It is at once apparent that the main variable term in the difference between the diurnal variations at each of these two pairs of stations is *semi-diurnal*, in accordance with the previous findings for h' and $h^{\text{max.}}$. There is also a large constant-difference term between the average diurnal values of $f^0_{F_2}$ at the three latitudes, as evidenced by the displacement of the graphs above the zero of $f^0_{F_2}$. The average values at Cape York, and to a lesser extent Brisbane, are clearly considerably higher than those at Canberra. A simple explanation of this phenomenon will emerge from the theory outlined below.

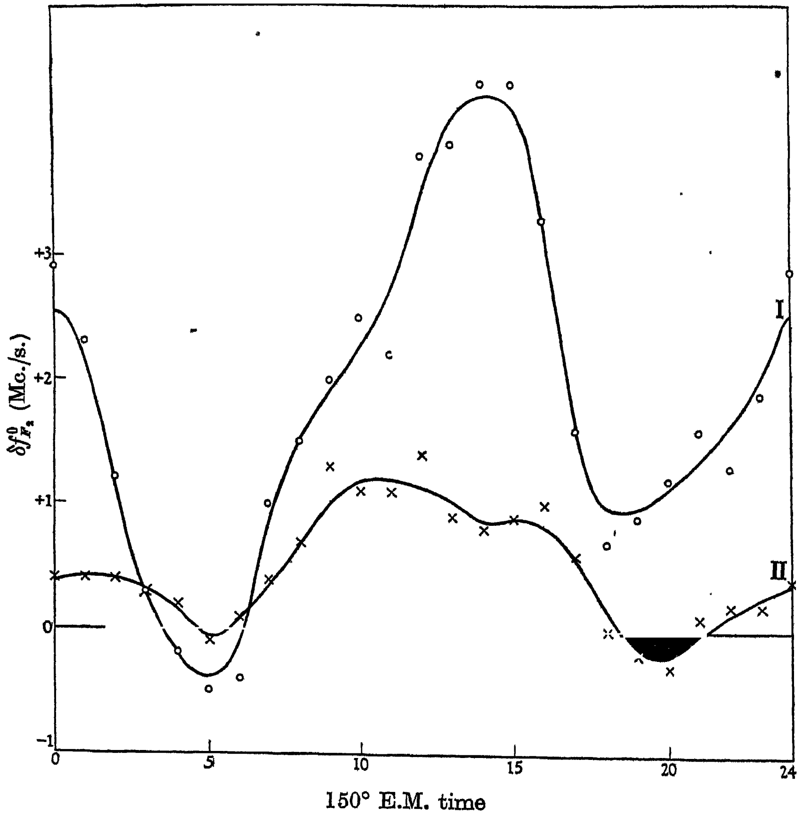


FIGURE 3. Difference between average diurnal variations of $f^0_{F_2}$ at Cape York, Brisbane and Canberra for March 1945. Curve I, Cape York-Canberra; curve II, Brisbane-Canberra.

3. MOTIONS OF IONS DUE TO HORIZONTAL WINDS

3·1. In figure 4, x, y, z form a system of right-handed axes. The earth's magnetic field H is directed along the z axis, which is inclined downwards, thus simulating conditions in the northern hemisphere. CA lies in the magnetic meridian, so that I is the magnetic dip. AB is the (horizontal) vector representing the air velocity c_0, so that $\pi + \phi$ is the direction of this wind with respect to magnetic north.

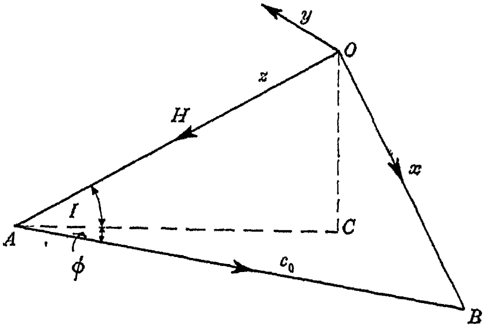


FIGURE 4. Diagram of right-handed axes system used in deduction of components of ionic motion.

In the absence of a magnetic field ions would be swept along with the wind with the velocity c_0. Such motion may be produced by the action of an average force F_0 due to the air. The application of the simplest considerations of kinetic theory to a lightly ionized gas gives

$$F_0 = m\nu c_0,$$

where m is the mass of an ion, ν is the mean collision frequency of an ion with the air molecules or atoms, neglecting constants of order unity which would arise from considerations of velocity distribution. $F_\parallel$ and $F_\perp$ are written for the components of F_0 along the z and x axes respectively. The presence of a magnetic field will not affect F_0 but will have a profound influence on the resulting ionic motion. If $\dot{x}'$, $\dot{y}'$, $\dot{z}'$ be the mean ionic velocities in the presence of H, then it may readily be shown (Chapman & Bartels 1940) that

$$\dot{x}' = \frac{F_\perp}{m\nu}\frac{\nu^2}{\nu^2+\omega^2} = c_\perp \frac{\nu^2}{\nu^2+\omega^2}, \tag{1}$$

$$\dot{y}' = -\frac{F_\perp}{eH}\frac{\omega^2}{\nu^2+\omega^2} = -c_\perp \frac{\nu\omega}{\nu^2+\omega^2}, \tag{2}$$

$$\dot{z}' = c_\parallel, \tag{3}$$

where $\omega = He/m$.

e is the ionic charge, and $c_\parallel$ and $c_\perp$ are the components of c_0 along the z and x axes respectively. Equations (1) and (3) express the well-known conclusion that diffusion of ions across a magnetic field is impeded, while that along the field is unaffected. Equation (2) expresses the 'dynamo' electric force, motion of ions of opposite signs being in this case in opposite directions owing to the presence of ω in the first power in the equation. In general, then, (2) will give rise to an electric current and also to a polarization of the medium.

3·2. Consider first the case where polarization is negligible. If v be the vertical component of ionic mass movement or 'drift', measured positively downwards, then by equations (3) and (1), and remembering that $c_{\parallel} = -c_0 \cos\phi \cos I$, then

$$v = c_{\parallel} \sin I + c_{\perp} \sin I \cot \angle OAB \frac{\nu^2}{\nu^2+\omega^2}$$

$$= -\tfrac{1}{2} c_0 \cos\phi \sin 2I \frac{\omega^2}{\nu^2+\omega^2}. \tag{4}$$

In the F_2 region $\nu \ll \omega$ for both electrons and ions, so that the vertical drift velocity is very nearly equal to the vertical component of the air velocity component $c_{\parallel}$. There is thus a vertical migration of ions of all kinds upwards or downwards according as $\phi \lesseqgtr |\tfrac{1}{2}\pi|$, *with a velocity of the same order of magnitude as the horizontal wind velocity*, and given exactly by equation (4).

3·3. Let it now be supposed that the polarization of the medium is such that the current due to $\dot{y}'$ is inhibited. Let the electric field due to this polarization be E_0, then for no current it follows that

$$\dot{y}'_0 = -\dot{y}',$$

or

$$\frac{E_0 e}{m\nu} \frac{\nu^2}{\nu^2+\omega^2} = c_{\perp} \frac{\nu\omega}{\nu^2+\omega^2},$$

or

$$E_0 e = m\omega c_{\perp}.$$

Due to E_0 there is now an additional x velocity $\dot{x}'_0$ such that

$$\dot{x}'_0 = +\frac{E_0 e}{eH} \frac{\omega^2}{\nu^2+\omega^2} = c_{\perp} \frac{\omega^2}{\nu^2+\omega^2}.$$

The total x velocity is now

$$\dot{x}' + \dot{x}'_0 = c_{\perp}\left(\frac{\nu^2}{\nu^2+\omega^2} + \frac{\omega^2}{\nu^2+\omega^2}\right) = c_{\perp}.$$

The ionic flow along the x axis is now unimpeded by the magnetic field, and on combining the flow along the x and z axes c_0 is obtained. *Thus when polarization prevents the flow of current due to the 'dynamo' electric force the ions flow in the air stream completely unaffected by the magnetic field, and no vertical drift occurs.*

3·4. In general both current flow and polarization will occur in the earth's atmosphere. In fact some polarization will usually be necessary in order to permit the current to flow in a closed circuit, as in the dynamo theory of magnetic variations. In such events the total vertical ionic drift may be deduced by combining v_1, that due directly to air motion, with v_2, that due to the electric polarization field E, which may be capable of specification when the boundary conditions are laid down. Ionic drift can only be produced by $E_\perp$, the component of E perpendicular to H. If $E_\perp$ is further resolved into two components $E'_\perp$ and $E''_\perp$ lying respectively in the horizontal plane and in the vertical plane containing H, then only $E'_\perp$ can produce vertical drift. If $E'_\perp e$ is written for $F_\perp$, and the axes rotated about z until that of x is horizontal and westerly, then application of equation (2) gives

$$\dot{y}' = -\frac{E'_\perp e}{eH}\frac{\omega^2}{\nu^2+\omega^2},$$

and so
$$v_2 = +\frac{E'_\perp}{H}\frac{\omega^2}{\nu^2+\omega^2}\cos I, \tag{5}$$

the disappearance of e from the equations indicating that $\dot{y}'$ (and v_2) are now drift velocities independent of ionic sign.

It is now possible to deduce the total vertical ionic drift, provided the horizontal air velocity and electrostatic field be known.

4. VERTICAL MOTIONS OF IONS DUE TO ATMOSPHERIC TIDES

4·1. It will be assumed that the horizontal winds associated with atmospheric tides can be derived from a series of velocity potentials, a typical term of which, ψ_n^σ, is a surface harmonic of the general form

$$\psi_n^\sigma = A_n^\sigma P_{n,\sigma}\sin\{\sigma(\lambda+t)-\alpha_n^\sigma\},$$

where A_n^σ is a constant factor expressing the amplitude of the harmonic of order σ and degree n, $P_{n,\sigma}$ is the corresponding associated Legendre function, λ is the longitude at the selected point, t is time referred to a fixed meridian and α_n^σ is a phase constant. v_1 is now derived in terms of ψ. Equation (4) shows that v_1 is proportional to the north component $-c_0\cos\phi$ of wind velocity, which may be expressed as $d\psi/a\,d\theta$, here θ is the co-latitude, and a the radius of the earth. I may also be expressed in terms of θ using the well-known relationship $\tan I = 2\cot\theta$. Equation (4) then becomes

$$v_1 = \frac{\sin 2\theta}{4-3\sin^2\theta}\frac{d\psi}{a\,d\theta}\frac{\omega^2}{\nu^2+\omega^2}.$$

In deriving v_2 it is assumed, in the first instance, that the conductivity, at the selected level, is uniform over the earth. In this case Schuster (1908) has shown that the horizontal electrostatic field is derivable from a potential S given by

$$S = +\frac{H\sin I}{\cos\theta}\frac{d\psi}{d\lambda}\Big/ n\,.\,n+1.$$

It follows in this case that

$$E'_\perp = +\frac{dS}{a\sin\theta\,d\lambda} = +\frac{2H\sin I}{a\sin 2\theta}\frac{d^2\psi}{d\lambda^2}\Big/ n\,.\,n+1,$$

and so, by equation (5),

$$\begin{aligned} v_2 &= +\frac{\sin 2I}{n(n+1)\,a\sin 2\theta}\frac{\omega^2}{\nu^2+\omega^2}\frac{d^2\psi}{d\lambda^2} \\ &= -\frac{2\sigma^2\psi}{n(n+1)\,a(4-3\sin^2\theta)}\frac{\omega^2}{\nu^2+\omega^2}. \end{aligned} \tag{5a}$$

Hence
$$v = v_1 + v_2$$
$$= \frac{\omega^2}{a(\nu^2+\omega^2)(4-3\sin^2\theta)}\left\{\sin 2\theta \frac{d\psi}{d\theta} - \frac{2\sigma^2}{n(n+1)}\psi\right\}.$$

Applying now the reduction formula
$$\frac{dP_{n,\sigma}}{d\theta} = \sigma \cot\theta P_{n,\sigma} - P_{n,\sigma+1},$$

and expressing ψ in its most general form, then
$$v = \frac{\omega^2}{a(\nu^2+\omega^2)(4-3\sin^2\theta)}$$
$$\times \sum_{\sigma=0}^{\infty}\sum_{n=1}^{\infty} A_n^\sigma\left\{2\sigma\left(\cos^2\theta - \frac{\sigma}{n\,.\,n+1}\right)P_{n,\sigma} - \sin 2\theta P_{n,\sigma+1}\right\}\sin(\sigma(\lambda+t)-\alpha_n^\sigma).$$

In the important case of the sectorial harmonics ($n=\sigma$), the second term within the brackets vanishes, since $P_{r,r+1} = 0$. The dependence of v on co-latitude θ is illustrated in figure 5*a* for the sectorial harmonics ψ_1^1 and ψ_2^2, and for the tesseral harmonics ψ_2^1 and ψ_3^2. In the figure is plotted the function $F(\theta)$ given by
$$(4-3\sin^2\theta)F(\theta) = \left\{2\frac{(n-\sigma)!}{(n+\sigma)!}\right\}^{\frac{1}{2}}\left\{2\sigma\left(\cos^2\theta - \frac{\sigma}{n\,.\,n+1}\right)P_{n,\sigma} - \sin 2\theta P_{n,\sigma+1}\right\}.$$

The Schmidt factor $\left\{2\frac{(n-\sigma)!}{(n+\sigma)!}\right\}^{\frac{1}{2}}$ serves roughly to normalize the coefficients of the various harmonics. It is seen from the figure that, in general, for each harmonic the sense of the vertical velocity changes in passing from high to low latitudes.

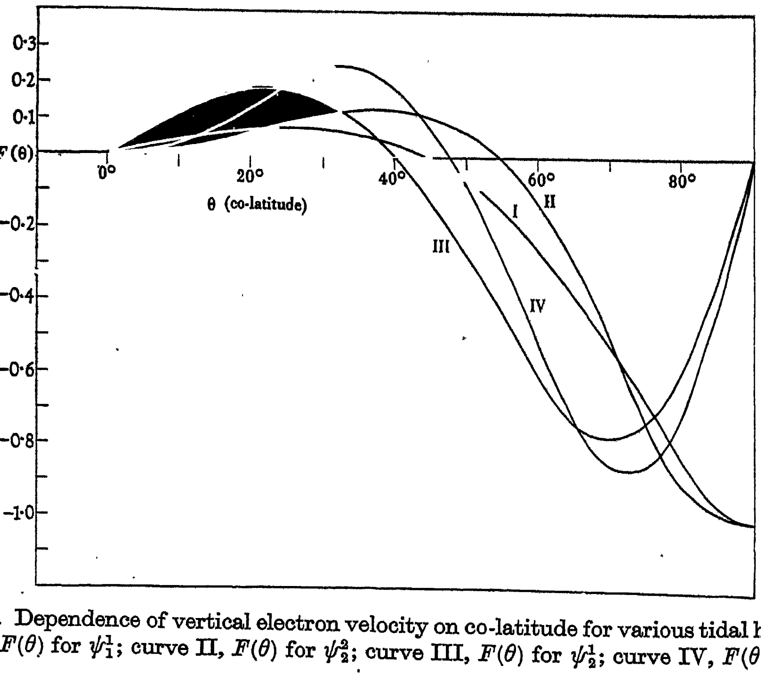


Figure 5*a*. Dependence of vertical electron velocity on co-latitude for various tidal harmonics. Curve I, $F(\theta)$ for ψ_1^1; curve II, $F(\theta)$ for ψ_2^2; curve III, $F(\theta)$ for ψ_2^1; curve IV, $F(\theta)$ for ψ_3^2.

4·2. When the non-uniformity of conductivity over the earth is taken into account the problem presents considerable analytical difficulties. Use may be made of certain results of Schuster (1908) in order to assess the importance of this effect. The latter considered a variation of the conductivity ρ given by

$$\rho = \rho_0\{1+\tfrac{1}{4}\pi\sin\delta\cos\theta+\tfrac{1}{4}\pi\cos\delta\sin\theta\cos(\lambda+t)\},$$

where δ is the solar declination.

It may then be shown, following Schuster, that

$$S = \frac{H\sin I}{\cos\theta}A^{\sigma}_{n}\sum_{\tau=-\infty}^{+\infty}k^{\tau}_{m}P_{m,\tau}\cos(\tau(\lambda+t)-\alpha^{\tau}_{m}).$$

The tidal velocity potential whose amplitude is A^{σ}_{m} now gives rise to an infinite series of harmonics in the polarization field. The amplitudes of these harmonics depend on the coefficients k^{τ}_{m}, some of which have been evaluated by Schuster for the two cases A^1_1 and A^2_2. The more important of these are as follows:

Case 1. A^1_1: $k^1_1 = \tfrac{1}{2}+\tfrac{1}{80}\gamma^2$, $k^2_2 = -\gamma/72$.

Case 2. A^2_2: $k^2_2 = \tfrac{1}{3}+\tfrac{2}{189}\gamma^2$, $k^1_3 = -\tfrac{1}{18}\gamma$, $k^2_3 = \tfrac{1}{45}\gamma^1$,

where $\gamma = \tfrac{1}{4}\pi\cos\delta$ and $\gamma^1 = \tfrac{1}{4}\pi\sin\delta$.

To compare the amplitudes of the vertical velocities due to the various harmonics $dS/d\lambda$ must be compared for each harmonic. Thus for the case of A^2_2 the ratio $(E'_{\perp})^1_3/(E'_{\perp})^2_2$ is found to be $5\pi(\cos^2\theta-\tfrac{1}{5})/98\sin^2\theta$, which is much less than unity save in the immediate vicinity of the poles. A similar result is obtained on comparing the other principal harmonics arising from non-uniform conductivity with the main term.

It appears therefore that such inferences as may be made regarding vertical motions of ions from the simple treatment given above in § 4·1 are not likely to be seriously affected by the assumption of uniform conductivity. This conclusion is not, of course, applicable to the dynamo theory of magnetic variations, which is ultimately concerned with horizontal electric *currents*, and not, as in this paper, with horizontal electric *fields*.

4·3. So far consideration has been given to the consequences of horizontal motions of ions across only the vertical component of H. East-west motion across the horizontal component of H must give rise to vertical electric forces whose effects will now be examined. Around a given parallel of latitude these forces will be directed upwards at some points and downwards at others, so that polarization may be expected to result in the completion of current circuits in the vertical east-west plane. In his classical paper Schuster drew attention to this possibility (1908, para. 8), but pointed out that the magnetic consequences at ground-level would be negligible, since the sum of the current flow at all levels above a given point due to this cause would be zero. Our concern, however, is with the consequences, at any selected level, of horizontal electric forces at that level. Clearly there is here a cause of vertical ionic motion whose magnitude must be examined.

Pekeris (1937) has shown that the amplitude of the tidal velocity changes markedly with height when account is taken of the vertical temperature gradient. There is also a change of phase of π at a certain level. Some unpublished work by G. Odgers seems to show that further changes of phase and amplitude occur when Pekeris's work is extended to the higher regions of the ionosphere above 100 km. For the purposes of the present analysis it will be assumed that ψ_n^σ contains a factor $\sin(by-\beta)$, where y is the height above an arbitrarily selected datum level, β is unknown, and b is of order of magnitude 100 km.$^{-1}$. The x axis horizontal, pointing positively to magnetic east, is chosen. The surface conductivity ρ_s is assumed uniform in the vertical current sheet to be considered. E_x, E_y are written for the dynamo electric forces in the x, y directions respectively, and E'_x, E'_y for the corresponding forces due to the polarization potential S. The current function in the sheet is represented by R. Then

$$E_x = 0, \quad E_y = -\frac{d\psi}{a\,d\lambda}H/\sin\theta\,\cos I$$

$$= -a^{-1}H/\sin\theta\,.\cos I A_n^\sigma P_{n,\sigma}\,\sigma\sin(by-\beta)\cos(\sigma\lambda-\alpha)$$

$$= B/\sin\theta\,.\sin(by-\beta)\cos(\sigma\lambda-\alpha), \quad \text{say},$$

$$E'_x = -\frac{\partial S}{\partial x}, \quad E'_y = -\frac{\partial S}{\partial y},$$

$$-\frac{\partial R}{\partial y} = -\rho_s\frac{\partial S}{\partial x}, \quad \frac{\partial R}{\partial x} = -\rho_s\frac{\partial S}{\partial y} - \rho_s B/\sin\theta\,.\cos(\sigma a^{-1}x-\alpha)\sin(by-\beta).$$

Eliminating S from the latter two equations then

$$\frac{\partial^2 R}{\partial x^2}+\frac{\partial^2 R}{\partial y^2} = -\rho_s B/\sin\theta\,.\sigma a^{-1}\sin(\sigma a^{-1}x-\alpha)\sin(by-\beta),$$

the relevant solution of which is obviously

$$R = a\rho_s B/\sin\theta\,.\sigma(a^2b^2+\sigma^2)^{-1}\sin(\sigma a^{-1}x-\alpha)\sin(by-\beta).$$

It follows that $E'_\perp$ (as formerly defined) is given by

$$E'_\perp = -\rho_s^{-1}\frac{\partial R}{\partial y} = -H/\sin\theta\,.\cos I\,\sigma^2 b(a^2b^2+\sigma^2)^{-1}\psi_n^\sigma\cos(by-\beta),$$

and so, by equation (5)

$$v_2 = -\frac{\sin\theta}{4-3\sin^2\theta}\frac{b}{a^2b^2+\sigma^2}\frac{\omega^2}{\nu^2+\omega^2}\sigma^2\psi_n^\sigma\cos(by-\beta).$$

To assess the magnitude of this velocity due to *vertical* polarization, compare it with that in equation (5*a*), which gives the corresponding velocity due to *horizontal* polarization. Neglecting terms of order unity it is seen that the ratio of the two velocities is $ab/(a^2b^2+\sigma^2)$. Now a is 6400 km. and b is of order 100 km.$^{-1}$, so that for all the lower order harmonics likely to be of importance, v_2 due to vertical polarization is small compared with that due to horizontal polarization, and will not be discussed further.

4·4. So far it has been assumed that the magnetic and geographic poles coincide. The removal of this assumption leads to the appearance of terms in v whose magnitudes depend on λ_0, the longitude measured from the meridian (68° W of Greenwich) which contains the earth's magnetic pole. The general problem has been attacked by Schuster (1908) and Chapman (1919) in developing the dynamo theory of magnetic variations. These authors' results will be applied to this problem for the special case of ψ_2^2, which tidal theory indicates as likely to be the most important term in the velocity potential. If the earth is regarded as a uniformly magnetized sphere, whose magnetic axis is inclined at an angle Φ to the geographical axis, then it follows from these authors' results (e.g. Schuster 1908, p. 171), that the additional term in $E_\perp$ is

$$E_\perp = \frac{1}{15a} A_2^2 C \tan\Phi \frac{d}{d\theta}\{(27P_{1,1}-2P_{3,1})\sin(2t'-\lambda_0-\alpha)+P_{3,3}\sin(2t'+\lambda_0-\alpha)\}$$

$$= a^{-1}C\tan\Phi A_2^2\cos\theta\{\sqrt{[(4-3\cos 2\theta)^2\cos^2\lambda_0+\sin^2\lambda_0]}\sin(2t'-\alpha+\epsilon), \qquad (6)$$

where C is a constant of the earth's field, and $\tan\epsilon = (3\cos 2\theta-4)^{-1}\tan\lambda_0$. The expression under the $\surd$ sign has its maximum value when $\lambda_0 = 0$ or π, and its minimum when $\lambda_0 = \frac{1}{2}\pi$ or $\frac{3}{2}\pi$. When $\lambda_0 = 0$ the additional term in $E_\perp$ becomes

$$-a^{-1}C\tan\Phi A_2^2\cos\theta(4-3\cos 2\theta)\sin(2t'-\alpha),$$

and so the additional term in v is

$$-a^{-1}H^{-1}C\cos I\tan\Phi A_2^2\frac{\omega^2}{\nu^2+\omega^2}\cos\theta(4-3\cos 2\theta)\sin(2t'-\alpha).$$

Now when $\lambda_0 = 0$ the well-known formula for the distribution of H round a magnetized sphere becomes

$$-\frac{C\cos I}{H} = \frac{2\sin(\theta-\Phi)}{4-3\sin^2(\theta-\Phi)},$$

so that the additional term in v is

$$a^{-1}A_2^2\tan\Phi\frac{\omega^2}{\nu^2+\omega^2}\frac{2\sin(\theta-\Phi)}{4-3\sin^2(\theta-\Phi)}\cos\theta(4-3\cos 2\theta)\sin(2t'-\alpha).$$

(This expression contains the additional terms in both v_1 and v_2, since equation (6) has been derived from Schuster's current function, which includes both the dynamo and the polarization field.) The function $F_1(\theta)$ given by

$$F_1(\theta) = 0{\cdot}058\frac{2\sin(\theta-\Phi)}{4-3\sin^2(\theta-\Phi)}\cos\theta(4-3\cos 2\theta)$$

is plotted in figure 5*b*. The factor 0·058 is the product of the Schmidt normalizing factor for $P_{2,2}$ and $\tan\Phi$, which has the value 0·202. It will be observed that $F_1(\theta)$ is unsymmetrical about both the geographic and the magnetic equators. It attains its greatest value south of both equators for the meridian $\lambda_0 = 0$.

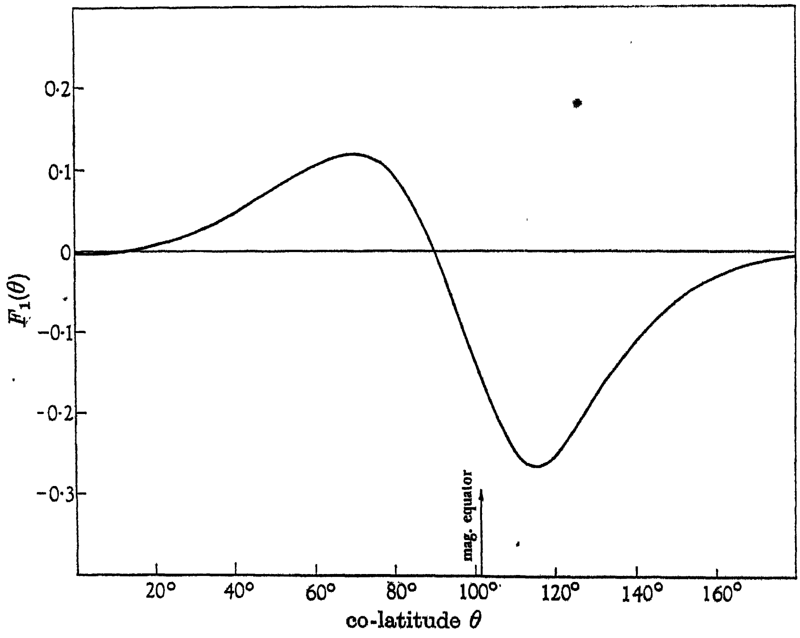


FIGURE 5*b*. Additional vertical velocity term due to magnetic asymmetry.

The term in v due to magnetic asymmetry may now be compared with the main term, for the case of ψ_2^2, by comparing the curve in figure 5*b* with curve II in figure 5*a*. Study of these curves shows that the asymmetric term is comparable with the main term over a range of latitudes from about 25 to 55°. It is concluded that over this range of latitude there will appear a 'longitude effect' in v, whose magnitude and phase will depend on longitude measured from 68° W of Greenwich.

5. THE INFLUENCE OF VELOCITY-HEIGHT GRADIENT

5·1. It remains to consider the gradient of v with height, and the effect of this on the form of an ionized region. The work of Pekeris (1937) and of Odgers (unpublished) shows that in the lower regions of the ionosphere the variation with height of tidal velocities (on which v ultimately depends) is not easily formulable. In these regions damping is negligible and tidal velocities are a very complicated function of the temperature-height gradient. There is little doubt, however, that at the height of the F_2 region damping due to viscosity and heat conduction are important terms in the equations. (Pekeris, indeed, concludes that at 200 km. damping is very large, and dominates the other terms, but it seems certain that this is an over-statement based on the assumption, no doubt for reasons of mathematical simplicity, that the average temperature of the ionosphere is only 210° K.) Both the kinematic viscosity and the thermometric conductivity, on which damping depends, are inversely proportional to the gas pressure, which decreases exponentially with height. It seems reasonable to assume then that, in the F_2 region, tidal velocities, and therefore v, decrease exponentially with height. Therefore

$$v = Ve^{-\gamma z},$$

where z is the 'reduced' height (Chapman & Bartels 1940, vol. 1, p. 506), measured in units of H_0 the 'height of the homogeneous atmosphere', and γ is a positive constant of order unity.

5·2. When account is taken of the vertical motion, the equation giving the electron density is

$$\frac{\partial N}{\partial t} = I'(z,t) - \alpha' N^2 + \frac{\partial (Nv)}{\partial z}, \tag{7}$$

where $I'(z, t)$ is the rate of ion production per c.c., and α' is the effective recombination coefficient after account is taken of attachment and detachment. The values of v necessary to make the last term important are examined first. Since N and v both may be expected to have, in general, an exponential variation with height, the order of magnitude of this term will be that of Nv, if v be measured in units of H_0. The ratio of the orders of magnitude of the two last terms in equation (7) is therefore $v/\alpha' N$. N is about 10^6/c.c., and the usually accepted values of α' are about 10^{-10}. (It must be remembered, however, that the method used for determining α' (Appleton 1937) ignores the influence of v and may give much too high values if the views here presented be accepted.) Vertical movement becomes important therefore if v is of the order $10^{-4}H_0$, or 400 cm./sec. The velocity of the wind associated with the solar tide is about 30 cm./sec. at ground-level, and calculations according to Pekeris's methods lead to amplifications of several hundred times at ionospheric levels. There is therefore little doubt that movement is an important and probably a dominant factor in determining the observed values of N.

5·3. In order to bring out the important consequences of movement in as simple a manner as possible night-time conditions ($I' = 0$) are considered, and recombination neglected.*

An initial distribution of ionization corresponding to a Chapman region is assumed, i.e.

$$N_0 = n_0 \exp \tfrac{1}{2}(1 - z - \sec\chi\, e^{-z}).$$

Under these limitations the solution of equation (7) becomes

$$N = n_0 \eta^{-(2\gamma+1)/\gamma} e^{\frac{1}{2}(1-z)} \exp\left(-\tfrac{1}{2}\sec\chi\, \eta^{-1/\gamma} e^{-z}\right),$$

---

* Dr J. C. Jaeger is making a special study of this equation with the ionizing term included. The author is much indebted to him for making available his preliminary results, and for valuable discussion.

where $\eta = (1+\gamma Vte^{-\gamma z})$. A maximum occurs in N at the height z_m given by

$$\sec\chi\, e^{-z_m}\,\eta_m-(\gamma+1)/\gamma-(2\gamma+1)\,\eta_m^{-1}+2\gamma = 0.$$

In figure 6 N is shown as a function of z, for $\chi = 90°$, for two values of γ (0·25 and 1·0), and for several positive values of β ($=\gamma Vt$), i.e. for v downwards. It will be observed that in all cases the value of N(max.) is less than N_0(max.) and becomes progressively smaller with lapse of time (β increasing). It is particularly to be remarked that while for small values of $\gamma(0{\cdot}25)$ z_m decreases steadily with time, *for larger values* ($\gamma =$ 0·5, 1·0) *the maximum rises*. (The curves for $\gamma = 0{\cdot}5$ are not reproduced, to avoid confusing the figure.) Figure 7 shows the same type of curves for upward velocities ($\beta - ve$). It will be observed in these cases that N(max.) increases with time and z_m rises in general. There is an obvious tendency for the region to develop a sharp gradient on the under-surface, a tendency which eventually must be balanced by diffusion.

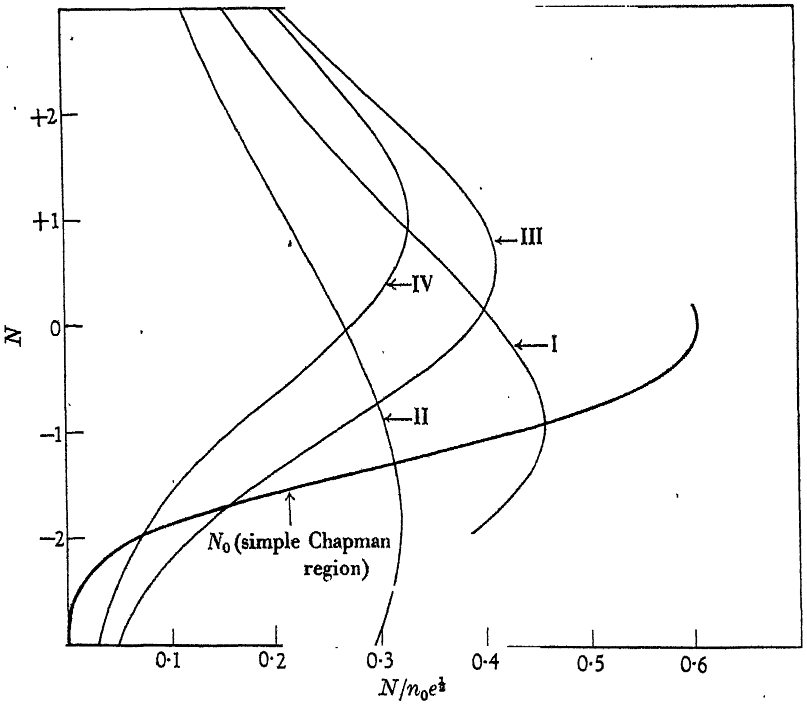


FIGURE 6. Effect of downward electron velocities on Chapman region. Curve I, $\gamma=0{\cdot}25$, $\beta=0{\cdot}25$; curve II, $\gamma=0{\cdot}25$, $\beta=0{\cdot}5$; curve III, $\gamma=1{\cdot}0$, $\beta=0{\cdot}5$; Curve IV, $\gamma=1{\cdot}0$, $\beta=1{\cdot}0$.

For small values of β (small velocities or small lapse of time) it is easy to show that equation (7) reduces to

$$z_m = \gamma Vt(2\gamma - 1/\gamma).$$

At this height N(max.) is then given by

$$N(\text{max.}) = n_0(1-\gamma Vt).$$

Thus in this special case, which it will appear in a later paper is of special importance for lunar tides, the movement of the maximum is in the opposite direction to that of V (for $\gamma > 0{\cdot}706$) and $N(\text{max.}) \gtrless N_0(\text{max.})$ according as $V \lesseqgtr 0$. This effect can be seen in figure 7, for the curve represented by the parameters $\gamma = 1$, $\beta = -0{\cdot}25$. It will be noticed that if in this case one is able experimentally to determine the values of z_m and $N(\text{max.})$ simultaneously, then one can evaluate both γ and V. It has been found possible to do this for the lunar tide which has been found in the F_2 region, but the results of this investigation must be deferred to a later paper.

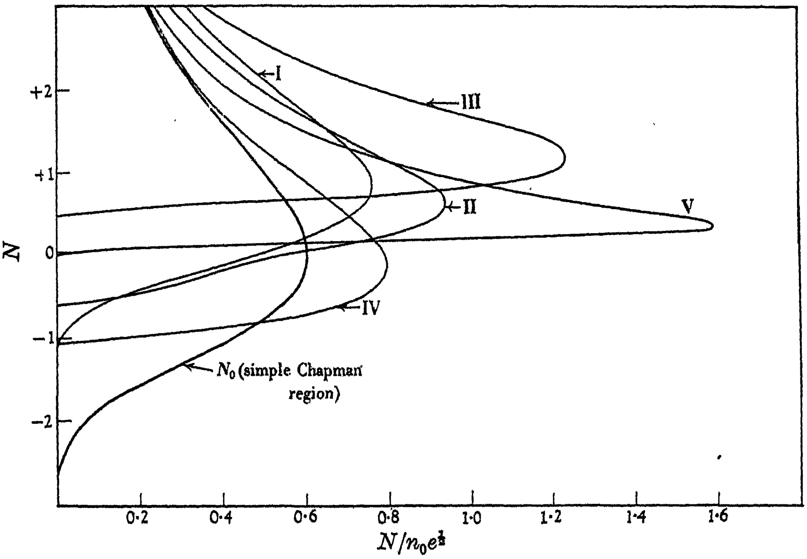


FIGURE 7. Effect of upward electron velocities on Chapman region. Curve I, $\gamma = 0{\cdot}25$, $\beta = -0{\cdot}25$; curve II, $\gamma = 0{\cdot}5$, $\beta = -0{\cdot}5$; curve III, $\gamma = 0{\cdot}5$, $\beta = -1{\cdot}0$; curve IV, $\gamma = 1{\cdot}0$, $\beta = -0{\cdot}25$; curve V, $\gamma = 1{\cdot}0$, $\beta = -1{\cdot}0$.

6. CONCLUSION

6·1. It is proposed in this section to review briefly some of the main features of F_2 region behaviour and the application of the above theory to their elucidation, leaving the detailed application of the theory to later papers.

First the value of α', the effective recombination coefficient in this region, is discussed. It has been a matter of considerable theoretical difficulty for some time that the measured value (c. 10^{-10}) is so much greater than the theoretical value for radiative recombination (c. 10^{-12}). On the other hand, there seems little evidence of a definite eclipse effect in this region, particularly when F_2 is well separated from F_1 at the time of eclipse. If the view advanced in this paper is accepted, the usual methods of measuring α' from observations of the diurnal change of $N(\text{max.})$ must be invalidated, and very little evidence indeed is left regarding the value of α'. In two papers, Woolley (1946) and Allen (1946) put forward evidence, the first from theoretical and the latter from observational reasoning, that the time of decay of F_2 ionization is of the order of a day and $\alpha' < 10^{-11}$.

Whatever the precise value of α' in F_2 may eventually prove to be, there is little doubt that it is very much less than that for F_1, which disappears as a separate region near sunset, and which shows a well-marked eclipse effect. The conclusion may safely be drawn that ionization in the F region which is moving upwards when formed by solar radiation will have a considerably longer life than similarly formed ionization moving downwards.

6·2. It is now well known that the ionization densities found in low latitudes in the F_2 region are much higher than can be accounted for by the simple Chapman theory. This fact is a natural consequence of the theory outlined in this paper. If Pekeris's conclusion that the phase of the solar tide in the ionosphere is opposite to that at the ground is accepted, then at latitudes above about 35° (the precise value depending on the 'longitude term' of § (4·4)) v is directed downwards between the hours of sunrise and noon (see figure 5a). Figure 6 shows that for such conditions N(max.) is reduced considerably below the Chapman value, and $h^{\text{max.}}$ rises, as is observed. A further cause of reduced N, which will be effective throughout the ensuing 24 hr., is the flow of freshly liberated electrons downwards into the 'sink' at lower levels where α' is high. In the afternoon electrons will flow upwards, leading to the afternoon concentration (figure 7) which is a well-marked feature of stations at these latitudes. The fact that at high latitudes the F_2 region is relatively free from this effect in winter, but shows it strongly in summer, suggests the presence of the seasonal harmonic ψ_3^2 (figure 5a) in intensity comparable with ψ_2^2. After sunset there should be a rapid drop of ionization until midnight (v downwards), followed by a slower drop (or even a rise in N(max.) if enough electrons persists at the lower levels) lasting till about sunrise. These effects also are well known observational phenomena.

At low latitudes from 0600 to 1200 hr. v is directed upwards and figure 7 shows that N(max.) will increase to values well above the Chapman value, and will rise steadily in height (after a small initial drop). Electrons are now being carried upwards on liberation and will therefore contribute to a high general level of N(max.) throughout the ensuing 24 hr. From 12 to 18 hr. electrons will move downwards again and $h^{\text{max.}}$ and N(max.) will fall, but from 18 to 24 hr. there may occur a further concentration leading to abnormally great N(max.). After midnight the values of N(max.) should commence to drop markedly until dawn. These phenomena too are well marked in the data from low-latitude stations. For example, the data for Cape York (lat. 11·0° S) used in figure 3 show a well-marked maximum in N(max.) at midnight.

Observatories very near to the magnetic equator show a slight change in the above-mentioned diurnal sequence. Here towards noon h'_{F_2} attains very high values (c. 400 km.) and N(max.) shows a well-marked minimum. Also, the diurnal variation of $h^{\text{max.}}$ is retarded in phase a few hours behind h'. It appears likely that this distinctive 'equatorial' behaviour is associated with the very great increase in v

(figure 5 *a*) which occurs in these latitudes. This must raise the ionization to abnormally great heights, as is observed. At these heights damping must be very great and, as is well known from the theory of damped oscillations, there will be a phase lag which may amount in the limit to $\frac{1}{2}\pi$. This would readily account for the phase lag in $h^{\text{max.}}$, and may also account for the noon 'bite-out' in N(max.). This point will be more fully discussed in a later paper.

6·3. Appleton & Weekes (1939) have found a semi-diurnal lunar tidal effect in the E region such that this region is 0·8 km. higher than average near the time of lunar transit. It has not so far been possible to interpret this result as due to vertical air motion associated with the lunar tide (Chapman & Bartels 1940). Any such interpretation encounters the grave difficulty that the phase of the observed motion is almost in exact opposition to that demanded by the dynamo theory and by Pekeris's calculations. The theory of vertical movements of ionization developed for the F_2 region in this paper cannot be directly applied to this problem, since there is no information regarding the height gradient of tidal velocities in the E region. It may be said, first, that the E region should have either a minimum or a maximum height at lunar transit, and so no theoretical impasse arises in the present state of our knowledge; secondly, it is a necessary consequence of our theory that for the small values of V associated with the lunar tide, there should be a phase difference of π between the vertical movements of ionized regions in high and low latitudes respectively. Appleton & Weekes's result refers to high latitudes (51° N). It has been found that the E region lunar tidal effect at Brisbane (lat. 27·5° S) shows the *minimum* height occurring slightly before lunar transit. This is almost exactly in phase opposition to that found in England and would appear to give strong support to our theory of the mechanism involved. This point will be fully discussed in the second paper of this series.

6·4. Finally, it is desirable to investigate briefly the reason why solar tides appear to exert such a strong influence in the F_2 and not in the E and F_1 regions. The most likely explanation of this fact appears to be that both I', the rate of ion production, and α' the effective recombination coefficient, are much higher in the latter two regions than in F_2. These two terms may be expected to dominate the tidal term in equation (7) so that these regions behave approximately like a simple Chapman region. (It should be noted, however, that there are small but definite anomalies in the global distribution of h'_E and h'_{F_1}, which may eventually prove to be of tidal origin.) Again it should be remembered that the vertical *gradient* of V is of decisive importance in determining the amount of region distortion. γ is large in the F_2 region because of damping. If the E and F_1 regions are situated in regions where this gradient is not great they will be little affected, particularly as regards N(max.).

This work has been carried out while the author was on an extended visit to the Commonwealth Observatory, Canberra, A.C.T. His thanks are due to the Commonwealth Astronomer, Dr R. v. d. R. Woolley for providing facilities and much helpful discussion and criticism; also to Dr G. W. Allen, G. Odgers, and Dr S. C. B. Gascoigne of the Observatory staff. The valuable contributions of Dr J. C. Jaeger of the Mathematics Department, University of Tasmania, have been mentioned above. Mr J. W. Reed assisted considerably with computations. The data used in figure 3 have been made available by the Australian Radio Propagation Committee. It is a pleasure to acknowledge the interest of the Chief Executive Officer of the Council for Scientific and Industrial Research, Sir David Rivett, F.R.S., who arranged the conditions necessary for the prosecution of this research.

REFERENCES

Allen 1946 *Terr. Mag. Atmos. Elect.* **51**, 1.
Appleton 1935 *Phys. Rev.* **47**, 89.
Appleton 1937 *Proc. Roy. Soc.* A, **162**, 451.
Appleton & Weekes 1939 *Proc. Roy. Soc.* A, **171**, 171.
Balfour Stewart 1882 *Encyclopaedia Britannica*, 9th ed.
Booker & Seaton 1939 *Trans. Wash. Meet. Ass. Terr. Mag. and Elect.* p. 503.
Chapman 1919 *Phil. Trans.* A, **218**, 1.
Chapman 1931 *Proc. Phys. Soc. Lond.* **43**, 26.
Chapman & Bartels 1940 *Geomagnetism.* Oxford: Clarendon Press.
Fleming 1939 *Terrestrial Magnetism and Electricity*, p. 38. New York: McGraw Hill.
Godfrey & Price 1937 *Proc. Roy. Soc.* A, **163**, 228.
Hulbert 1929 *Phys. Rev.* **34**, 1167.
Hulbert 1934 *Phys. Rev.* **46**, 822.
Martyn 1945 *Nature*, **155**, 363.
Martyn & Pulley 1936 *Proc. Roy. Soc.* A, **154**, 455.
Pekeris 1937 *Proc. Roy. Soc.* A, **158**, 650.
Schuster 1908 *Phil. Trans.* A, **208**, 163.
Woolley 1946 *Proc. Roy. Soc.* A, **187**, 102.

Temperature changes in the crystal structure of barium titanium oxide

By Helen D. Megaw, *Girton College, Cambridge*

(*Communicated by J. D. Bernal, F.R.S.—Received* 15 *March* 1946)

[Plate 5]

Barium titanium oxide, which is tetragonal at room temperature, changes about 120° C to a cubic structure. This change has been followed in detail by means of X-ray powder photographs taken in a 19 cm. powder camera at intervals of a few degrees over a range covering the transition point.

The unit cell, which contains the formula number of atoms, retains its identity throughout the transition, and the atomic parameters are unaltered. The change is simply in the axial lengths, and these vary continuously with the temperature, though not linearly, the variation becoming more rapid near the transition point. While the linear expansion coefficients along and perpendicular to the tetrad axis are large and of opposite sign, the volume expansion coefficient is small and positive. There is no discontinuous change either of linear spacing or of volume detectable at the transition point, but there is a sharp discontinuity in the linear expansion coefficients, and a marked increase in the volume expansion coefficient which is probably, though not certainly, discontinuous.

The transition suggests a typical λ-point change. The specific heat has not been determined, but the thermal expansion curve has the characteristic λ shape. Co-existence of cubic and tetragonal structures, in proportions depending on the temperature, occurs over a range of some degrees near the transition point, and is attributed to the effect of local stresses in facilitating or hindering a change between two structures whose energy difference is very small in this temperature range.

Below room temperature, observations made down to −183° C suggest that the structure may have a second transition point somewhere below this and become cubic again, the change being of the same nature as that at 120° C.

It is argued that the room-temperature structure can only be explained by the existence of directed bonds, and that the breaking of these bonds with increasing temperature is responsible for the 120° C transition. The low-temperature transition is explained by postulating a more complete set of bonds, probably an octahedral complex, which partially breaks down at this temperature to give the square formation observed in the room-temperature structure. The possible nature of the directed bonds is discussed qualitatively. The condition which makes possible the formation of such bonds is likely to be the abnormal volume available to the Ti atom, which is due to the effect of the large Ba ion in forcing apart the oxygen lattice. The directed bond system will only contribute a small part to the attractive energy of the lattice, which is mainly ionic in character. The hypothesis that directed bonds exist, whatever their origin, is used for a tentative explanation of anomalous variations of intensity of the X-ray lines observed at temperatures near the transition point.

Introduction

The crystal structure of barium titanium oxide (barium titanate) has recently attracted much attention. Early work by Barth (1925) showed that it belongs to the perovskite type. This has a simple cubic lattice, with the formula number of atoms per cell, the atomic parameters being as follows:

Ti	000,
Ba	$\frac{1}{2}\frac{1}{2}\frac{1}{2}$,
O	$\frac{1}{2}00$, $0\frac{1}{2}0$, $00\frac{1}{2}$.

Figure 1 shows this structure. It consists of a network of TiO_6 octahedra linked through each corner with their neighbours, the Ba ions being accommodated in the large central holes of the network, where they are surrounded by twelve oxygens.

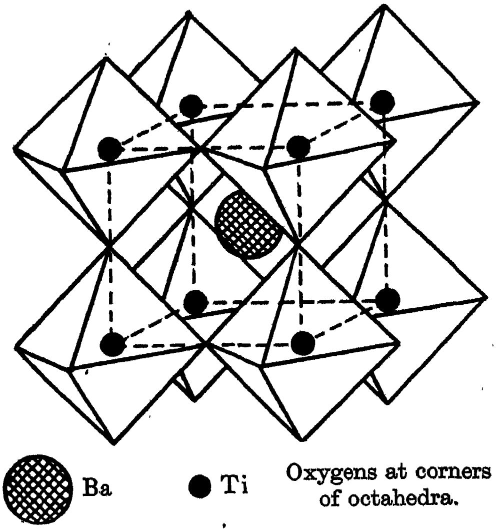


FIGURE 1. Structure of $BaTiO_3$. The unit cell is shown by the dotted line.

Recently it has been shown (Megaw 1945; Rooksby 1945) that $BaTiO_3$ is not strictly cubic at room temperature, but tetragonal; the atomic parameters remain unchanged,* but the whole unit cell is uniformly stretched about 1 % in the direction of one of the cube edges, which becomes the *c* axis.

Measurements made in the Material Research Laboratory of Philips Lamps, Ltd., of the variation of dielectric constant with temperature (Rushman & Strivens, 1946) showed a curve of the type recently described by Wul & Goldman (1945*a*; Wul 1945), with a conspicuous peak at about 120° C, and this immediately suggested that a change of structure might be involved. Investigations were accordingly made using a high-temperature camera in which the specimen could be examined over this temperature range.

It was found that the tetragonal form did in fact change its structure at about this temperature, inverting to the ideal cubic form described above; further, that the transition involved no discontinuous change either of spacing or of axial ratio. The detailed course of the temperature variation of spacing and of axial ratio was followed, and will be described below. In addition, there were certain peculiarities observed in the intensities of the X-ray lines, which will be discussed.

---

* The atomic parameters are fixed by the symmetry requirements if the lattice has a centre of symmetry. If there is no such centre, displacement along the tetrad axis becomes possible, and the space-group is lowered from D_{4h}^1–$P4/mmm$ to C_{4v}^1–$P4\,mm$ or S_4^1–$P\bar{4}$. This possibility is considered later in the text.

The curve of the variation of dielectric constant with temperature shows also a subsidiary peak at about 0° C. This suggested that it would be of interest to study the structure below 0° C. No accurately controllable low-temperature camera was available, and the experiments done were of the nature of a preliminary exploration only. While no change of structure was found at the lowest temperature reached (about $-183°$ C), there were changes of axial ratio which suggested that at a still lower temperature the structure might revert to cubic; and there were also intensity anomalies resembling those associated with the high-temperature transition.

The small amount of silica present in the material, as described below, whether it forms a solid solution or occurs as a separate phase, does not seem to have any important effect on the temperature changes, but further confirmation of this point would be desirable.

1. HIGH-TEMPERATURE TRANSITION

Experimental

The high-temperature camera used was that designed by A. J. C. Wilson (1941). It is a 19 cm. powder camera, with an oven surrounding the specimen, which can be kept at an accurately controlled temperature while the photograph is being taken.

The specimen was made from a sample of $BaTiO_3$ containing 1·5 % silica, fired at 1300° C to form a sintered disk. From this a piece was cut, ground and etched to give a rod about 2 cm. long and 1 mm. in diameter. This was mounted on the axis of the camera and rotated during exposure.

Copper radiation from a demountable tube was used, with a nickel filter. Exposures were about 1½ hr.

Photographs were taken at 20° intervals from 100° to 200° C, and at 2° intervals in the neighbourhood of the transition point, as well as at room temperature. The temperature of the specimen could be controlled to within ½° C, and explorations with a thermocouple showed that temperature differences in the air space surrounding the specimen did not exceed this.

The photographs were measured with a comparator reading to 0·01 mm., and a modified Bradley-Jay extrapolation (Megaw 1946*b*) was carried out to determine the cell-dimensions and axial ratio.

Interpretation of photographs

Photographs taken at different temperatures are shown in figure 2, plate 5. The high-angle lines are most important for the evaluation of cell-constants, and the others may be ignored. The cubic structure gives a set of doublets formed by reflexion from the same plane of the two wave-lengths α_1 and α_2. In the tetragonal structure each doublet is split into a set of partly overlapping doublets. The components of the highest-angle group, that with $h^2+k^2+l^2=26$, are tabulated below, with their Bragg θ-values. (The theoretical intensity values refer only to relative intensities of lines of the same group on the same photograph.)

TABLE 1

structure	indices and radiation	theoretical intensity ratio (allowing for overlaps)	$\theta°$
cubic at 200° C	$(431+501)\ \alpha_1$	2	78·25
	$(431+501)\ \alpha_2$	1	78·95
tetragonal at 20° C	$105\alpha_1$	1	77·03
	$314\alpha_1+105\alpha_2$	2·5	77·84
	$413\alpha_1+314\alpha_2$	3	78·57
	$^*(431+501+510)\ \alpha_1+413\alpha_2$	5	79·43
	$^*(431+501+510)\ \alpha_2$	2	80·24

* *Note.* 431 and 501 have geometrically identical spacings; 510 differs very slightly, but is not resolved as a separate line.

Results

It can be seen from figure 2, plate 5, how the lines of the tetragonal structure close together with increasing temperature, and finally coalesce in the cubic structure. This gives a measure of the decreasing axial ratio, which finally becomes unity. At the same time there are changes in the relative intensities of the lines. These will be discussed later; but here it may be noted that they make accurate determinations of the axial ratio near the transition point very difficult. Lines characteristic of the tetragonal structure have been observed up to 118° C, and may possibly be present at 120° C. Above this, only cubic lines can be detected.

The cell dimensions at the different temperatures are given in table 2. The last column gives the cell edge of those structures which are actually cubic, and the corresponding mean spacing, or cube-root of the volume, of the tetragonal structures.

TABLE 2

temperature °C	a kX	c kX	c/a	$\sqrt[3]{\text{vol.}}$ kX
20	3·9865	4·0254	1·0098 ± 0·0001	3·9996 ± 0·0004
90	3·9920	4·0192	1·0068 ± 0·0002	4·0011 ± 0·0005
100	3·9926	4·0162	1·0059 ± 0·0002	4·0005 ± 0·0008
116	3·9962	4·0122	1·0040 ± 0·0005	4·0015 ± 0·0008
117	3·9962	4·0102	1·0035 ± 0·0010	4·0010 ± 0·0010
117·5	3·9967	4·0087	1·0030 ± 0·0010	4·0007 ± 0·0010
118	3·9975	4·0083	1·0027 ± 0·0010	4·0011 ± 0·0005
120	—	—	1·0000 (+0·0020)	4·0011 ± 0·0005
122	—	—	1	4·0011 ± 0·0005
124	—	—	1	4·0012 ± 0·0005
126	—	—	1	4·0014 ± 0·0005
128	—	—	1	4·0010 ± 0·0007
130	—	—	1	4·0011 ± 0·0005
140	—	—	1	4·0016 ± 0·0005
160	—	—	1	4·0028 ± 0·0005
180	—	—	1	4·0037 ± 0·0005
200	—	—	1	4·0040 ± 0·0005

These spacings have been plotted against the temperature in figure 3. The non-linear variation even below 116° C is apparent. Above this, between 116° and 130° C, the exact course of the curve is more uncertain. The full line represents its most probable course, which is not very different from its lower possible limit; its upper limit is given by the dotted line. Above 130° C the change of spacing is linear with temperature, within experimental error.

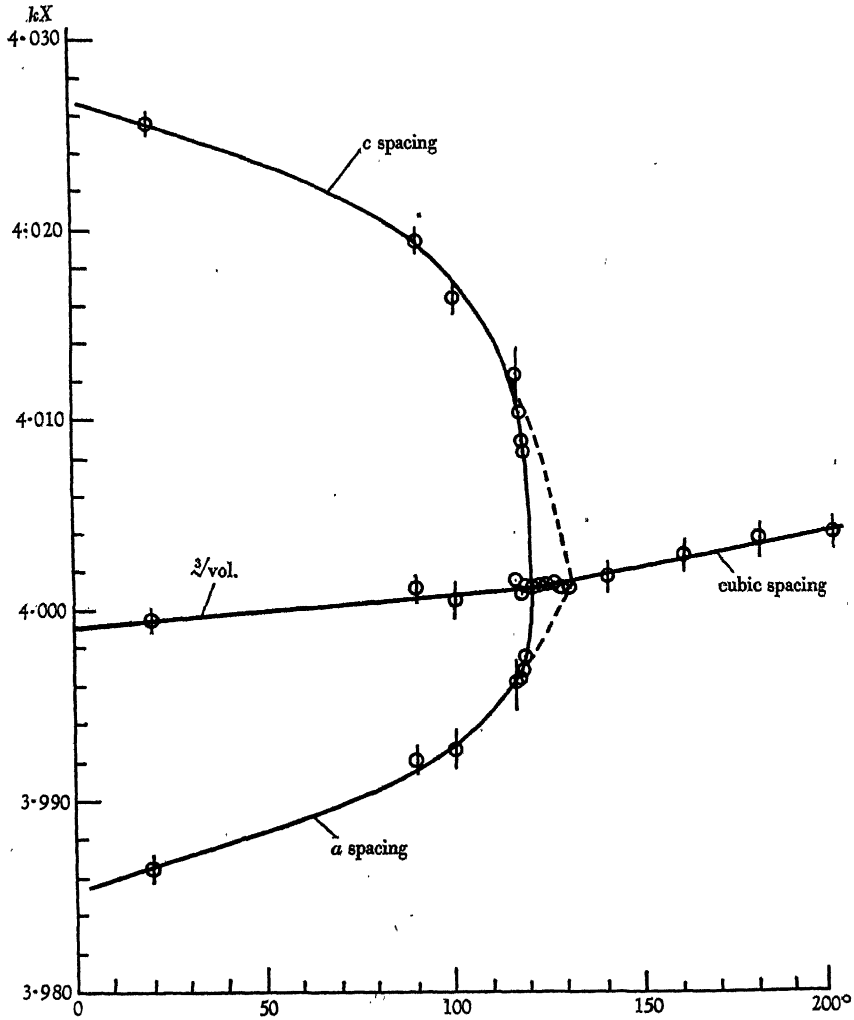


FIGURE 3. Temperature variation of cell dimensions of $BaTiO_3$. (The vertical lines through the points represent the limits of experimental error; those for observations near the transition point have, however, been omitted for the sake of clarity.)

The mean spacing of the tetragonal structure has been plotted against temperature in the same figure. It is noteworthy that the slope of the line is appreciably different from that of the cubic structure above 130° C, though of the same order of magnitude. In other words, the volume expansion coefficient, a constant for each structure over the temperature range studied, changes with the change of structure. Whether this occurs discontinuously at the transition point or gradually through intermediate values cannot be distinguished with the accuracy available. The values of the mean linear expansion coefficients, deduced from the slopes, are given in table 3.

TABLE 3

structure	temperature range °C	mean linear expansion coefficient × 10^6
tetragonal	20–120	$3{\cdot}5 \pm 1{\cdot}5$
cubic	130–200	10 ± 3

The linear thermal expansion coefficients in the a and c directions have been calculated from the slopes of the corresponding curves in figure 3, and plotted against the temperature in figure 4. These are necessarily very sensitive to the particular way in which the spacing/temperature curve is drawn through the experimental points, and can only be relied on as far as their general trend is concerned. They do, however, show at a glance how the coefficients are nearly constant at room temperature and increase rapidly in numerical value towards the transition temperature. Whether the region about 120° C should be occupied by a discontinuity or by a sharp peak cannot be determined on the available evidence; it depends on the exact shape selected for the spacing/temperature curve between its possible limits. The fact that the c expansion, which is negative in sign, shows the same sharp increase in numerical value as the a expansion, which is positive, rather suggests that there is a discontinuity in both at the transition temperature.

The axial ratio is plotted against temperature in figure 5. Here again the most probable curve has been drawn with a full line, the upper limit with a dotted line.

Intensity anomalies

In a perfect lattice, the relative intensities are determined by the geometry of the lattice, the atomic parameters, and the atomic form-factors or f-values of the atoms. For $BaTiO_3$ with its atoms in the special positions quoted above there are no atomic parameters which can vary with temperature. The f-values, which include the effect of thermal vibrations, are of course temperature-dependent; but changes in these, even relative changes as between the different atoms, will affect all the lines of a single group in the same way and to much the same extent, except in so far as the changes are anisotropic—an important exception which must be considered more fully later. The correction factor for polarization which has to be applied to the observed intensities may be neglected, since this varies continuously with θ, and the differences of θ within a group are small. Thus, unless anisotropic thermal vibrations are involved, the relative intensities within a group are

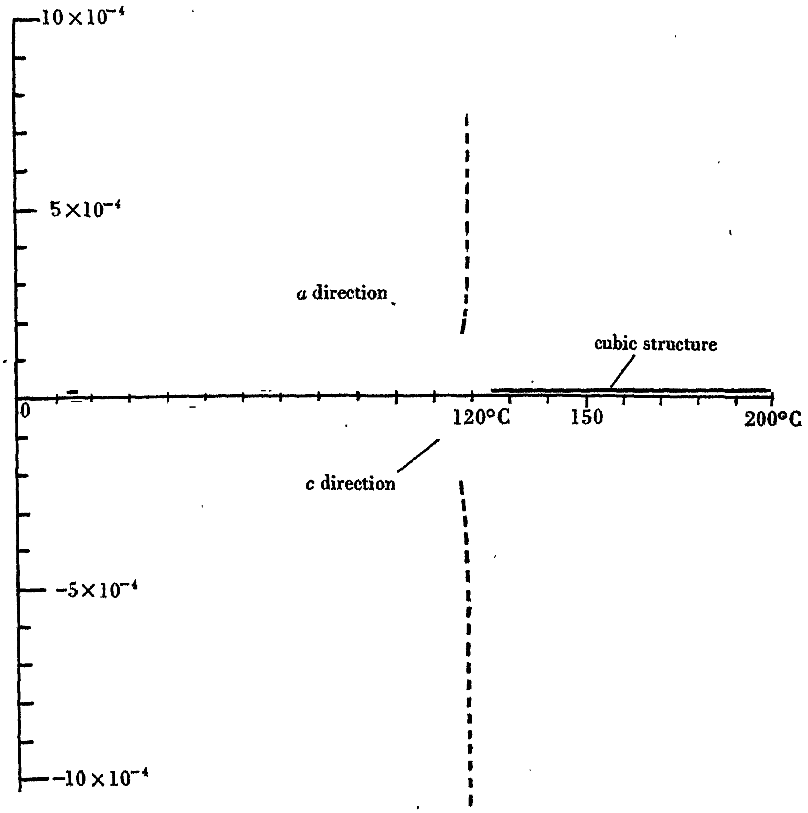


FIGURE 4. Linear expansion coefficients of $BaTiO_3$.

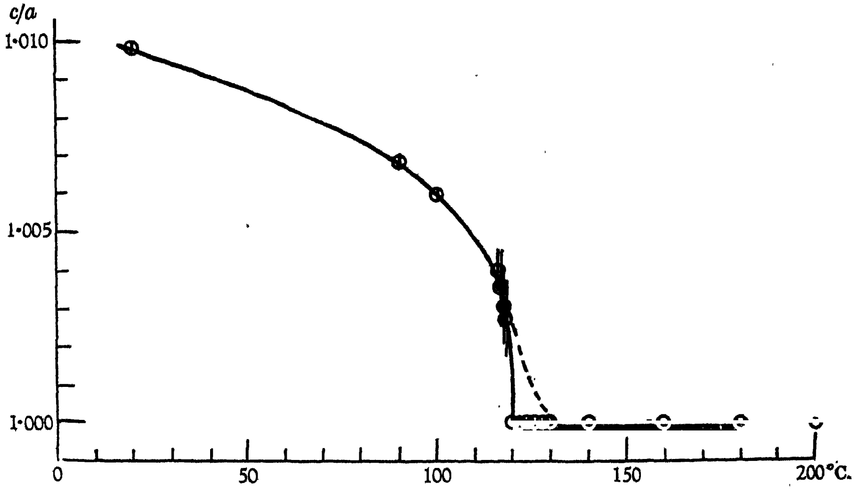


FIGURE 5. Temperature variation of axial ratio of $BaTiO_3$. (The vertical lines through the points represent the limits of experimental error.)

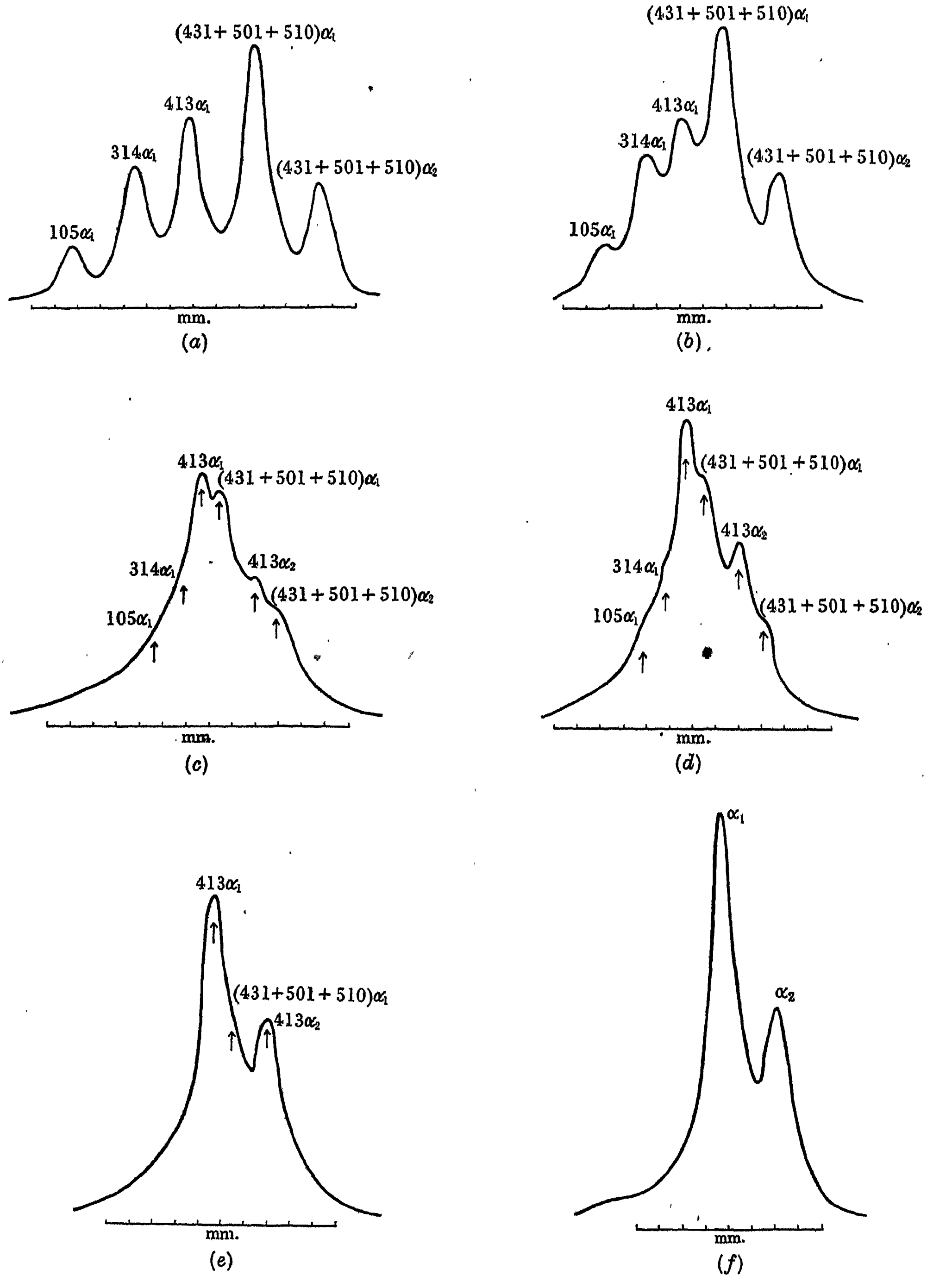


FIGURE 6. Photometer curves of line $N = 26$ at different temperatures. (a) 20°C, (b) 100°C, (c) 116°C, (d) 117°C, (e) 118°C, (f) 120°C.

practically independent of the temperature and of the axial ratio. In table 1 they are given for the high-angle lines which are of interest. They will be referred to as the 'normal intensities'.

Intensities were observed visually in the first instance. The observed intensities at room temperature agreed with the calculated, but at higher temperatures abnormalities were noticed, which fell into groups as follows:

(*a*) Lines of high l index became progressively weaker relative to other lines with the same $h^2+k^2+l^2$ when the transition temperature was approached. This appeared first, and was most marked, with line 105, but could also be seen with 314 and with 224.

(*b*) At temperatures within a few degrees of the transition, line 413 began to assume prominence, and increased in strength relative to the lines on either side of it till the transition temperature was reached.

(*c*) Near the transition point there was also a very marked increase in the background in the immediate neighbourhood of the lines.

Photometric measurements of the photographs were made, and confirmed these observations. The curves are shown in figure 6*a–f*, after subtraction of background; in drawing these, the ordinates were chosen to give all the same area, for convenient comparison, as the experimental intensity scale is arbitrary and not comparable between different photographs.

The decrease of distance between the peaks with increase of temperature is obvious and accounts for some of the diminution of importance of individual peaks. There is, however, a very striking increase of intensity of the 413 peak in the 116° C photograph, where it actually exceeds the 431 + 501 + 510 peak, hitherto strongest. This process is carried progressively further at 117° and 118° C.

To show up the same points more emphatically, a set of theoretical intensity curves was constructed for a normal tetragonal lattice with different axial ratios, and each matched against the experimental curve with the same axial ratio. For constructing such a theoretical curve, it was assumed to consist of a set of overlapping $\alpha_1\alpha_2$ doublets, each of the shape of the doublet at 200° C, where the structure is undoubtedly cubic. (Theoretically this doublet should be rather less sharp than at lower temperatures, but the difference is likely to be negligible for present purposes.) The ordinates of these doublets, each placed at its appropriate angle, were added to give the complete tetragonal curve. Figure 7 shows the results at 20°, 100° and 116° C. At 20° C the fit is good; the slightly greater sharpness of the experimental curve on the right-hand side probably comes within the limits of error. At 100° C the fit is reasonable, but the experimental 105 peak is lower than the theoretical, the experimental 413 peak higher. At 116° C the experimental 413 peak is very much too high, and the 105 and 314 peaks are smoothed out against a high general intensity covering their expected positions.

Such a change in the relative intensities is geometrically impossible in a perfect lattice, and it requires for its explanation either lattice imperfections of some kind or the co-existence of two structures in varying proportions. The two effects, enhancement of the 413 line and weakening of lines of high l, must be discussed separately.

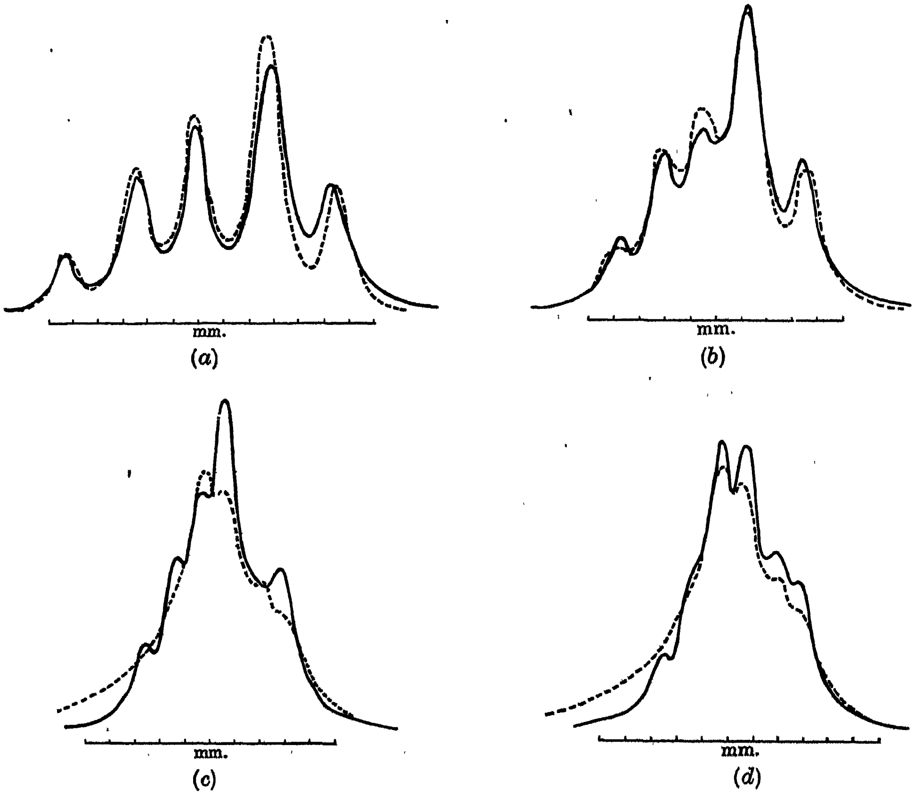


FIGURE 7. Comparison of calculated (——) and observed (– – – –) intensity curves. (*a*) 20° C, (*b*) 100° C, (*c*) 116° C, calculated for all the material tetragonal, (*d*) 116° C, calculated for 25 % of the material cubic, 75 % tetragonal.

Intensity of the 413 *peak*

While imperfections (using the term here to include thermal vibrations) may alter the relative intensities of different planes, the magnitude of the effect for any given plane will in general lie between the extreme values for the 001 and 100 planes, and in no case can be very much greater than for one of these. But the experimental fact here shows that 413 increases very greatly relative to both 105 (lying nearly parallel to 001) and 431 (nearly perpendicular to 001). Hence one must fall back on the explanation that a second structure has appeared, giving a strong line in the same position as the tetragonal 413. This condition would be satisfied by a cubic structure of the same cell-volume as the tetragonal, since their 413 spacings differ by less than 0·01 %. Other lines of this cubic structure would coincide with the corresponding tetragonal lines, which at the temperatures concerned are in any case unresolved single lines. Hence the observed intensity changes indicate the co-existence of a cubic and a tetragonal structure of the same cell-volume (or very nearly the same) over a temperature range of a few degrees near the 'transition point', which thus loses its exact meaning. There is in effect a *transition range.*

On this assumption, a theoretical intensity curve was constructed for 116° C, taking 25 % of the material as cubic and 75 % as tetragonal. This curve, shown in figure 7 (*d*), is a much better approximation to the experimental curve than is the pure tetragonal. At 118° C, a reasonable fit is obtained with a curve constructed for 67 % of the material cubic.

As quantitative estimates of the proportions of cubic and tetragonal material these are of course very rough, but they do illustrate the point that the proportion of cubic increases with the temperature.

That the effect is reversible, and not merely due to slow changes produced when the specimen is held at the high temperature, is shown by the fact that photographs taken at 117° and 117·5° C, at a time subsequent to all the others, fit into intermediate positions between the 116° and 118° C photographs.

Intensity of the 105 *peak*

In the above section, the working conclusion was reached that a cubic structure appears at some temperature below 120° C and gives a line overlapping with the tetragonal 413 line. This leaves for consideration the relative changes of the three remaining peaks, 105, 314 and 431 + 501 + 510.

Any imperfections affecting the distribution of scattering matter in the *c* direction will have a maximum effect on reflexion from the 001 plane and zero effect on planes perpendicular to this. This type of imperfection will therefore affect 105, 314 and 431 in decreasing order of magnitude. This is in accordance with the experimental results, where the weakening of 105 is most marked while 314 is also noticeably weakened. In the next group of lines, weakening of 224 relative to 422 is also observable.

There is thus evidence of changes affecting the *c* but not the *a* direction as the crystal approaches the transition range. They occur to an appreciable extent at temperatures below those where any noticeable amount of cubic material is formed. Possible mechanisms for them will be discussed later.

2. LOW TEMPERATURE CHANGES

Experimental

There was no camera ready available for low-temperature work, and circumstances limited what could be done to a quick general exploration.

A simple back-reflexion camera was used for most of the work. It was adapted so that a cylindrical film-holder, of about 6 cm. diameter, could be mounted round the specimen; this enabled a complete powder pattern to be obtained. No precautions were taken for centring the specimen on the axis of the camera, and thus the photographs were not suitable for measurement; but this was irrelevant to the main point at issue, as visual examination would suffice to tell whether the low-temperature pattern was essentially the same as that at room temperature.

For more detailed examination of the high-angle lines, a flat film was used in the back-reflexion holder, about 5 cm. from the specimen. A longer distance would have been preferable for increased accuracy, but would have prolonged the exposure times inconveniently.

The film was wrapped in a black paper envelope. In front of it was placed a film which had been uniformly exposed to light and developed. This formed a thin silver screen which absorbed most of the incoherent secondary radiation scattered by the titanium atoms, while only increasing the exposure time by a factor of about 2.

Two different types of specimen were made, both using finely powdered material from a fired disk. In one, the powder was packed into a thin celluloid tube; in the other, it was mixed with gum tragacanth, moistened, and rolled into a rod. Both kinds were roughly 5 mm. long and from 0·5 to 1 mm. in diameter. They were stuck, either with gum tragacanth paste or by moistening the celluloid with acetone, to a copper wire, which was coiled in a spiral about a knob in the bottom of a small Dewar cup, and secured in place by the same adhesive. The cup was mounted on the rotating axis of the camera so that the X-ray beam just cleared its rim. The specimen could be centred on the axis of rotation by manipulating the copper wire. The arrangement is shown in figure 8.

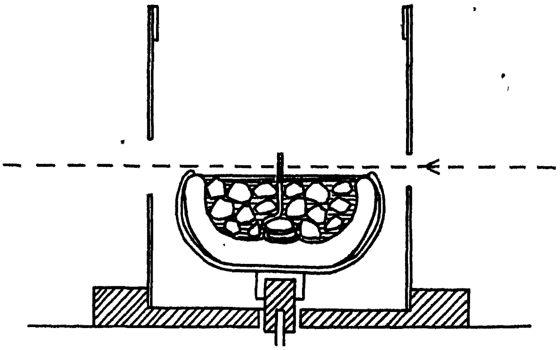


Figure 8. Experimental arrangement for low temperature photographs (showing cylindrical film holder).

The low temperatures used were the fixed points given by solid carbon dioxide (−78° C) and liquid oxygen (−183° C). Several different methods of achieving these were tried.

For −78° C, small lumps of solid CO_2 were fitted into the cup, which was filled with either methyl alcohol or acetone; the former was used with the celluloid tube, the latter with the gum-tragacanth rod. In both cases the specimen softened and bent slightly, but was not seriously displaced. The formation of ice crystals (hoar-frost) on the rim of the cup and on the specimen was prevented by washing these surfaces at intervals with cold liquid from the cup, applied with a small paint-brush.

The part of the specimen projecting above the cup will, of course, be at a temperature rather above −78° C. However, the ready formation of ice on it, if it is left unwashed, indicates that it is well below 0° C; that it is reasonably constant throughout the exposure in the part of the specimen irradiated may be deduced from the internal evidence of the photographs, namely the sharpness of the individual spots. In all these photographs a stationary specimen was used.

For liquid air temperatures, one photograph was taken with a stationary specimen, the cup being kept replenished with liquid air. In another photograph, the specimen was rotated, and a continuous stream of liquid air allowed to trickle down it from a funnel with a narrowed stem. Only back-reflexion photographs were taken with liquid air.

Results

Photographs taken on the cylindrical film at room temperature and with solid CO_2 showed essentially the same structure, giving the normal pattern for $BaTiO_3$. Close inspection, however, revealed small differences in the relative separations and intensities of high-angle lines, notably the group of lines with $h^2+k^2+l^2=26$. This group was examined in more detail with the back-reflexion camera.

In these photographs it was not possible to identify the separate lines of the group with absolute certainty. The difficulty is not wholly due to the rather primitive technique, but is partly inherent in the material. Low-temperature photographs with stationary specimens show individual spots about as sharp as those at room temperature (showing that variation of temperature during the exposure may be neglected) but the background of smaller spots is drawn out into a woolly band indicating a large amount of strain inhomogeneity. The same effect produces a blurring together of the lines in the photograph with the rotating specimen, in marked contrast to their good definition at room temperature. It resembles the blurring noticed just below 120° C, though it is still more pronounced, and it is attributable to the same cause. Typical photographs are shown in figure 9, plate 5.

In spite of this lack of clarity, two points stand out with certainty:

(i) In all the low-temperature photographs the axial ratio is markedly *less* than at room temperature.

(ii) At CO_2 temperatures the intensity relations of the lines are abnormal.

Measurement of the photographs gives the following estimate of the axial ratios:

With solid CO_2	$c/a = 1{\cdot}0085 \pm 0{\cdot}0010$.
With liquid air	$c/a = 1{\cdot}0030 \pm 0{\cdot}0020$.

The intensity relationships with solid CO_2 may most probably be accounted for by assuming that two effects are occurring simultaneously, as in the region of the high-temperature transition: (*a*) the co-existence of a cubic and a tetragonal structure, (*b*) a weakening of certain planes of the tetragonal structure relative to the rest, in this case the planes which lie most nearly *parallel* to the tetrad axis. A similar interpretation seems compatible with the liquid air photograph, though the blurring renders it much less certain.

The decrease of axial ratio suggests the existence of a low-temperature transition 'point' below which the structure reverts to cubic. The anomalous intensities suggest that this transition is of the same nature as the 120° change, though extended over a greater temperature range. Figure 10 shows the variation of axial ratio with temperature from −200° to +200° C, a dotted line being used where its course is only roughly known.

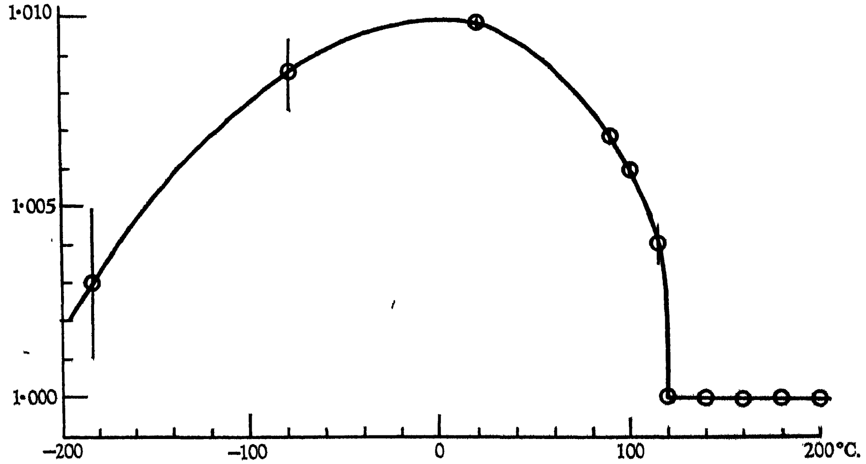


FIGURE 10. Variation of axial ratio with temperature. (The vertical lines through the points represent the limits of experimental error.)

DISCUSSION

The transition as a λ-point change

The changes occurring at the transition point of $BaTiO_3$ are obviously co-operative phenomena; the shape of the curves in figures 3 and 5 is that characteristic of a thermodynamic change 'of the second order'. This phrase was originally introduced (Ehrenfest 1933) to describe transitions where the changes of energy and volume were zero but the specific heat and thermal expansion showed discontinuities or anomalous increases. Whether this definition can be strictly upheld is now in question, and Ubbelohde & Woodward (1945*a*) would prefer to use the expression 'continuous changes'. They are often called 'λ-point phenomena' after the characteristic λ-shaped curves for the specific heat and the thermal expansion. They include the order-disorder changes in alloys, ferromagnetism, the transitions in ammonium chloride, methane, Rochelle salt, potassium di-hydrogen phosphate, and the α-β transition in quartz. Originally they were believed to occur without change of structure, but recent work (Ubbelohde & Woodward 1945*a*) has shown the change of structure accompanying the appearance of abnormal physical properties in Rochelle salt and KH_2PO_4. A comparison of $BaTiO_3$ with these two compounds is likely to be particularly illuminating, since it, like them, must be classed as a ferroelectric. In some of the compounds mentioned, for example, methane, NH_4Cl, and Rochelle salt, the transitions have been attributed to the setting in of rotation of certain molecules or ions, but this is by no means conclusively proved for any of them, and can certainly not be true for KH_2PO_4, quartz, or $BaTiO_3$.

The present work brings forward no experimental evidence on the variation of the specific heat of $BaTiO_3$ with temperature, but it is said by Wul & Goldman (1945*b*) to show 'an abrupt variation of the specific heat at a temperature of about 125° C'.* (On the other hand there is evidence from the work of Bantle (1942) on KD_2PO_4, another ferroelectric, that the specific heat anomaly may be structure-sensitive, and depend on the existence of sufficiently large Weiss domains.) The thermal expansion curves in figure 4 have the characteristic λ shape. There is no detectable volume discontinuity, a fact which brings it within the original definition of a second-order change (which would exclude, for example, that in NH_4Cl). In this it closely resembles the α-β transition in quartz, where X-ray examination of the spacings (Jay 1933) showed curves very similar to those of figure 3, except that for quartz both expansions are positive and the high-symmetry form is not cubic but uniaxial. In quartz, as in $BaTiO_3$, the linkage of the oxygen polyhedra remains unaltered throughout the transition, but in quartz they are able to tilt slightly from their high-symmetry positions, a possibility which in $BaTiO_3$ is incompatible with the retention of the small unit cell actually observed.

The double transition suggested for $BaTiO_3$ is paralleled by that observed in Rochelle salt. In it the high-symmetry form (here orthorhombic) exists above 24·5° and below −20° C, and the monoclinic structure in the intervening range is derived from the orthorhombic by a very slight distortion (Ubbelohde & Woodward 1946). As in $BaTiO_3$ (Megaw 1946*a*), the low-symmetry 'crystal' is an intimate twin of small regions oriented in certain specified directions relative to the original high-symmetry single crystal; and the change from single crystal to twin is reversible. The same effect has been studied in KH_2PO_4 (de Quervain 1944; Ubbelohde & Woodward 1945*a*, *b*), where the high-symmetry form is tetragonal. In both these the abnormal electrical properties are again associated with the low-symmetry range.

The co-existence of two forms over a finite temperature range has been observed in NH_4Cl, and the mechanism of the process has been very elegantly explained by

---

* (*Note added in proof.*) A recent paper by Wul (1946*b*) quotes measurements by Sokolov showing a maximum for the thermal capacity in the same region as for the dielectric constant. The same paper gives a very interesting thermodynamic treatment of the dielectric properties, though not directly linking it with the nature of the structural change.

Dinichert & Weigle (1944; Dinichert 1944). They suppose that, as the temperature rises, small regions of the high-temperature form which appear locally do not break away from the parent lattice, but produce local strains in their surroundings, which raise the transition temperature for the strained material. The different thermal expansions of the two phases then bring about the hysteresis which is characteristic of the change. Such hysteresis has not been observed in $BaTiO_3$, but the experiments were not designed to detect it.

In $BaTiO_3$, a change involving expansion in the a direction and contraction in the c direction will be facilitated by anisotropic local stresses where compressive forces in the c direction predominate. Crystals subject to such forces will invert to cubic at a lower temperature than those where the stresses are isotropic. Since, in a polycrystalline mass, there will be varying local stresses on the different crystallites, the structure change in the material as a whole will take place over a range of temperature, and will be reversible (though possibly with hysteresis) as long as the texture remains unchanged by recrystallization or crystal growth. Evidence of this reversibility has been obtained for $(Ba, Sr)TiO_3$ (Megaw 1946*a*), and less definitely for $BaTiO_3$ (unpublished observations). Calculations of the differential stresses required to produce a given change in the transition point of a crystallite have not been carried out, but, since the difference of volume between the two structures in the neighbourhood of the transition point is so small, the equilibrium temperature is likely to be very sensitive to such stresses. It is interesting in this connexion to notice that a dependence of dielectric constant on pressure has been observed by Wul & Vereshchagin (1945; Wul 1946*b*).

It is only possible to guess at a mechanism which would explain the λ-point. I am indebted to Dr Eisenschitz for the suggestion that one possibility is the assumption of two quantum states for the titanium atom, which do not combine, and which are degenerate for cubic symmetry but split up for tetragonal; to allow for the attainment of equilibrium between the two phases, it must be postulated that combination between the two states can occur if the crystal symmetry is lower, as it may be momentarily and locally owing to thermal vibrations. The atomic transitions will in this case be quite independent of one another, and the co-operative character of the change as a whole is only achieved by the accompanying alteration in atomic positions; in this it differs from ferromagnetism, where the atomic transitions are themselves co-operative.

Another possible mechanism which might give a λ-point is the existence of two equilibrium positions (one of them perhaps metastable) for one of the atoms. It is necessary before discussing this to consider a little more fully the evidence for locating the Ti atoms at the 000 positions. Figure 11 shows a section of the structure in the (100) plane at height $x = 0$. It has always been assumed that the structure has a centre of symmetry, as drawn in the figure; but if the centre were absent, displacement of atoms along the c axis would be permissible. It has not been possible so far to obtain conclusive evidence from piezo- or pyro-electric tests. The agreement of calculated and (visually) observed intensities in the room-temperature structure shows that the displacements, at least those of the cations, are not large, but does not rule them out completely. The abnormal steric conditions in $BaTiO_3$, which will be discussed more fully in the next section, allow small displacements of the atoms lying on the tetrad axis to be made without bringing them closer together than the sum of the Goldschmidt radii. Displacements with a statistical distribution of plus and minus signs might not give rise to a noticeable piezo-electric effect, and yet would not require a larger unit cell.

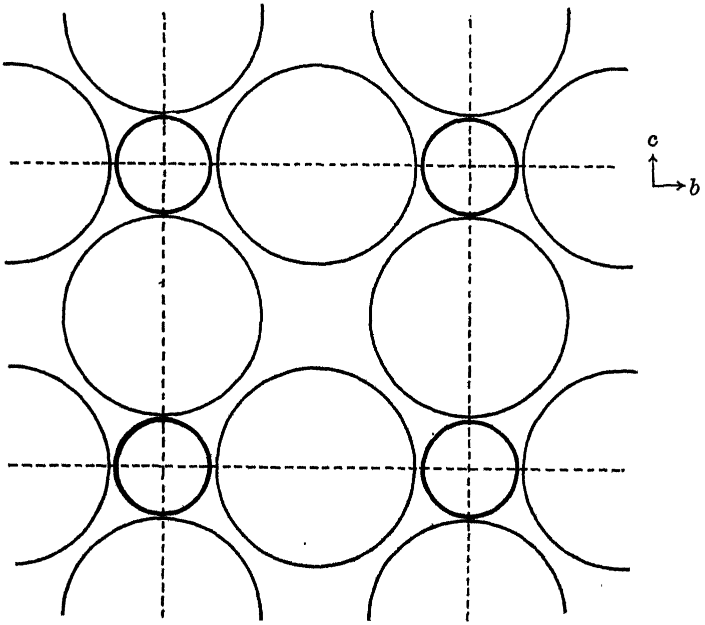


FIGURE 11. Section of structure of $BaTiO_3$ in plane (100) at height $x=0$. The small circles represent Ti ions, the large circles O ions, drawn to scale using the Goldschmidt radii; the dotted lines outline the unit cell.

The Ti-O distances in the c direction are 2·01 kX at room temperature, as compared with 1·96 kX for the sum of the Goldschmidt radii. Thus either the Ti or the O atom may be displaced by as much as 0·05 kX towards one of its neighbours; but at the same time the distance from its opposite neighbour is increased, so that the move may result in a gain in energy of the lattice as a whole. However, there may be some displaced position which is metastable with respect to the symmetrical position, and the existence of two discrete energy levels could be used to account for the λ-curve. It is not necessary to assume that any large number of the metastable positions are occupied at room temperature. This suggestion of a mechanism is only very tentative, but it seems compatible with the general properties of the structure.

The tetragonal structure

So far, no explanation of why the structure is tetragonal has been advanced. It is generally accepted that barium titanium oxide is an isodesmic compound (Evans 1938), since the forces throughout the lattice are essentially of the same kind; the electrostatic bond strengths are all less than unity, having the values 2/12 and 4/6 respectively for the Ba-O and Ti-O coordination polyhedra. It is most naturally considered as an ionic structure, since it satisfies all the usual criteria for ionic structures, with one exception: the difficulty arises in explaining its tetragonal symmetry, since that symmetry is intrinsic in the environment of each atom and is not merely caused by the geometrical building together of the co-ordination polyhedra.

The departure from cubic symmetry cannot be explained by purely steric considerations. In all cubic compounds of the type $A^{2+}B^{4+}O_3$ the cell-dimensions are determined by the network of linked $B^{4+}O_6$ octahedra, the large cation accommodating itself in the space between the octahedra. If this cation is rather too small for the hole, as in $CaTiO_3$, it causes a collapse of the oxygen framework of such a kind that the octahedra retain their dimensions and linkage but not their symmetrical orientation (Naray-Szabo 1943). $BaTiO_3$ is the unique example of a compound with a cation just too *large* for the available hole (Megaw 1946*b*). The oxygens round the Ti^{4+} ion are forced apart, leaving abnormally large Ti-O distances, 1·99 and 2·01 kX in the *a* and *c* directions respectively, as compared with 1·96 kX for the sum of the Goldschmidt radii. The Ti ion can 'rattle' in its oxygen octahedron. On purely steric grounds one might expect either the retention of a cubic lattice with all the distances equally too large, or the formation of a tetragonal lattice in which some of the distances, say those in the (001) plane, remained normal; or possibly the occurrence of some quite different modification where by means of lowered symmetry or a larger cell the normal distances were achieved; but not a structure in which the distances are *all* abnormally large but different in the different directions.

Since it is thus impossible to account for the existing structure by means of purely central electrostatic forces, one is driven to postulate that there are directed bonds determining the uniaxial symmetry. Accepting Pauling's view that bonds may be partly covalent and partly ionic, it is only necessary to assume a small amount of this covalent character superimposed on a mainly ionic set of bonds in order to explain the phenomena observed.

Since the cell-edge *a* in the room-temperature structure is shorter than *c*, the bonds must lie in the (001) plane. It is natural to assume that they have a square formation in this plane.

Evidence for directed bonds

The only atom to which the directed bonds can be attributed with any plausibility is the titanium. If the compound is strictly stoichiometric, we are concerned with Ti^{IV}. There is evidence that this forms octahedral bonds in the compounds

TiS_2, $TiSe_2$, $TiTe_2$. Pauling & Huggins originally (1934) considered these bonds to be $3d^2 4s 4p^3$, but later (1937, 1941) changed their opinion and concluded that they are $4s 4p^3 4d^2$, i.e. that they make use of the unstable d orbitals of the valence shell itself. Huggins states that these are more polar than the $3d^2 4s 4p^3$ bonds, and should predominate for bonds connecting a metal atom to very electro-negative atoms. These $4s 4p^3 4d^2$ bonds have properties intermediate between truly ionic and truly covalent bonds; their relationship to the ionic makes more plausible their appearance in such a compound as $BaTiO_3$.

The Ti-O distance which should be found with such bonds is given by the sum of the Pauling covalent radii as 2·02kX. Now, as pointed out above, it is an experimental fact that the Ti occupies an abnormally large volume for an ionic structure, the Ti-O distances being 2·01 and 1·99kX as compared with 1·96kX for the interionic distance normally to be expected. The bond orbitals are very sensitive to distance; and it seems probable that this abnormally large volume may favour the formation of a covalent bond system which is associated with longer Ti-O distances than the ionic. Because of this same sensitiveness, small changes in the interatomic distances due to thermal expansion may suffice to bring about the changes in the bond system which are needed to account for the change of structure.

There is no previous evidence whatever for the occurrence of square bonds with Ti, though they are of course well known for the transition elements. The square complex for Ti^{IV} may only be made possible by the very abnormal volume available to the atom.

On the other hand, it is possible that $BaTiO_3$ is not stoichiometric, but contains a small number of Ti^{3+} with corresponding vacant oxygen sites. It has not been possible to analyse the material with the necessary accuracy to test this. The large Ti volume would favour the occurrence of Ti^{3+}, which has a radius of 0·69A as compared with 0·64 for Ti^{4+}.

Nothing is known about the covalent bond system of Ti^{III}. However, a different kind of bond postulated by Klemm (1939) to occur between cations with incomplete outer shells, attributable to anti-parallel spins, has been invoked by Ehrlich (1939) to explain the magnetic susceptibility of TiO_2 on reduction. These bonds are not formed by Ti^{4+}, but as soon as reduction has given rise to some Ti^{3+} bond formation is possible. If there are very few Ti^{3+} ions, they will be so widely separated that the number of bonds is extremely small; as reduction increases the number of Ti^{3+}'s, the proportion of bonds increases and the proportion of 'free' Ti^{3+}'s drops. This accounts satisfactorily for the variation of paramagnetic susceptibility with the composition. The amount of reduction required to produce an appreciable number of bonds, while only a few per cent, is accompanied by very drastic colour changes.

If there is free Ti^{3+} in normal $BaTiO_3$ it should be shown by its paramagnetism, but so far no study of this is available. The colour of the material, which is typically pale yellow, does not suggest that there is very much reduction.

Suggested explanation of the transition

Without making any detailed assumptions about the origin of the directed bonds, it is possible to use the idea to explain the temperature changes of structure. A system is postulated of directed bonds superimposed on a mainly ionic structure, and only making a small contribution to the lattice energy; it is further postulated that these directed bonds will tend to weaken or disappear at high temperatures. To fit the experimental observations, it is assumed that at low temperatures the direction of the bonds is octahedral, and that the breakdown of the system takes place in two steps, giving as an intermediate stage a square bond complex.

Consider first the high-temperature transition, which has been the more fully examined. The directed bonds in the room-temperature structure lie entirely in the (001) plane, and as the temperature rises they break. There is no discontinuous change in lattice energy, and hence this disappearance of the directed bonds means a redistribution of the attractive potential, the loss in the part attributable to directed bonds being compensated by a gain in the ionic part, which is effective in all directions and thus tends to reduce the c spacing. This accounts for the abnormal linear expansions in the two directions, while the volume expansion remains roughly normal. It also suggests an explanation for the change in mean linear expansion coefficient at the transition point: the same temperature-increase of energy can bring about a greater increase of interatomic distance when acting against the purely ionic forces above the transition point than when it has to overcome the stronger, partly covalent, forces below it.

At low temperatures the directed bonds coincide with the ionic, both being satisfied by a cubic structure. The low-temperature transition is brought about by the breaking of those directed bonds which lie in the c direction.

Origin of anomalous intensities

Consider now the anomalous intensity changes accompanying the transitions. For the high-temperature transition, this was shown to involve changes in the distribution of scattering matter in the c direction but not in the a direction; while for the low-temperature transition it was concluded tentatively that the reverse was the case.

Displacements of either Ti or O atoms of the kind suggested earlier in the paper would bring about intensity changes in the right direction but of insufficient magnitude. For example, displacement of the Ti atoms by 0·05 A (the full amount allowed in the room-temperature structure by the Goldschmidt radii) would only reduce the intensity of the 105 line by about 5 %; and the distance available for such displacement becomes rapidly less with temperature, just over the region where the intensity differences increase most markedly.

There are four other possible explanations of the weakening of reflexions of high l.

(1) The amplitude of thermal vibrations parallel to the *c*-axis becomes very large.

(2) There is a *random* displacement in the *c* direction of individual atoms from their lattice positions.

(3) The single crystals may break up into small crystallites or 'islands' which are plate-like parallel to (001), their thickness perpendicular to this being small enough to lead to diffraction broadening.

(4) Individual crystals may show a range of *c* spacings brought about by local stresses, with consequent broadening of the 105 line. Since the stresses are likely to be roughly isotropic, this must imply that the Young's modulus in the *c* direction is very much smaller than in the *a* direction (a not unreasonable supposition for a material where the two thermal expansions are of opposite sign).

These four mechanisms are not entirely distinct. The last three differ from one another chiefly in the size of the regions throughout which they require the *c* displacements to be of the same kind; even in (2) there must be some sort of local order. For individual crystals which are large enough to give discrete spots in photographs with a stationary specimen, they should, in their extreme forms, be distinguishable experimentally by the character of the distortion of spots on the 105 line at a temperature just below the transition point. However, in practice the inhomogeneity in texture of the material masks this. All three effects may be operative, but to a different degree in each crystallite, depending in its size and the local stresses.

It is possible to put forward a crude picture to show why the random displacements postulated under (2) should be in the *c* direction at the upper transition point. Suppose that there is originally a square network of *a* bonds, and that one such bond is broken. The repulsive forces tend to make the atoms move farther apart, but they cannot move *in* the plane because of the rigidity of the square network which remains. However, there is only a very small component of the directed bond preventing their movement out of the plane, and as such movement is favoured by the increase in ionic attractive force in the *c* direction it accordingly takes place.

The occurrence of thermal vibrations of large amplitude would lead to much the same physical picture. These imply a weakened restoring force in the *c* direction, a result which would be expected from the breaking of the *c*-bonds. Since there is not room for large enough amplitudes to be achieved by individual atoms independently of their neighbours, the instantaneous description of the crystal would show some degree of order, as would also occur with static displacements. So far as the present work goes, both these possibilities remain open.

As increasingly large numbers of bonds are broken, a progressive crumpling-up of the (001) plane may ensue, in which the atoms are displaced, or vibrate, perpendicular to the plane with only local order; or the displacements within a given region may all be of the same kind, resulting in an island of tetragonal structure as under (3).

The existence of such islands has been shown by Bradley (1940) for certain Fe-Cu-Ni alloys cooled too quickly for equilibrium. In these, the equilibrium structure consists of two face-centred cubic phases with slightly different dimensions. The partly quenched material contains islands sharing a common (001) plane with the parent structure, but having different axial ratios, $1{\cdot}009_5$ and 0·992 in a typical example. These islands are platelets about 100 A thick in the *c* direction which have not yet broken away from the parent lattice to assume their own characteristic *a*-spacing. There is a similarity here to the mechanism postulated by Dinichert & Weigle (1944) for NH_4Cl. Owing to the thinness of these plates in the *c* direction, they show diffraction broadening for lines of high *l*, the 004 line being washed out completely against the background. The experimental evidence very closely resembles that for $BaTiO_3$, except that in the latter case there is only one tetragonal lattice involved, with an axial ratio greater than unity. There is, however, one important difference in the conditions: in the Fe-Cu-Ni alloy, the effect is essentially one of incomplete equilibrium, the partial separation having been 'frozen in' by subsequent cooling, whereas in $BaTiO_3$ as in NH_4Cl there appears to be at least quasi-equilibrium.

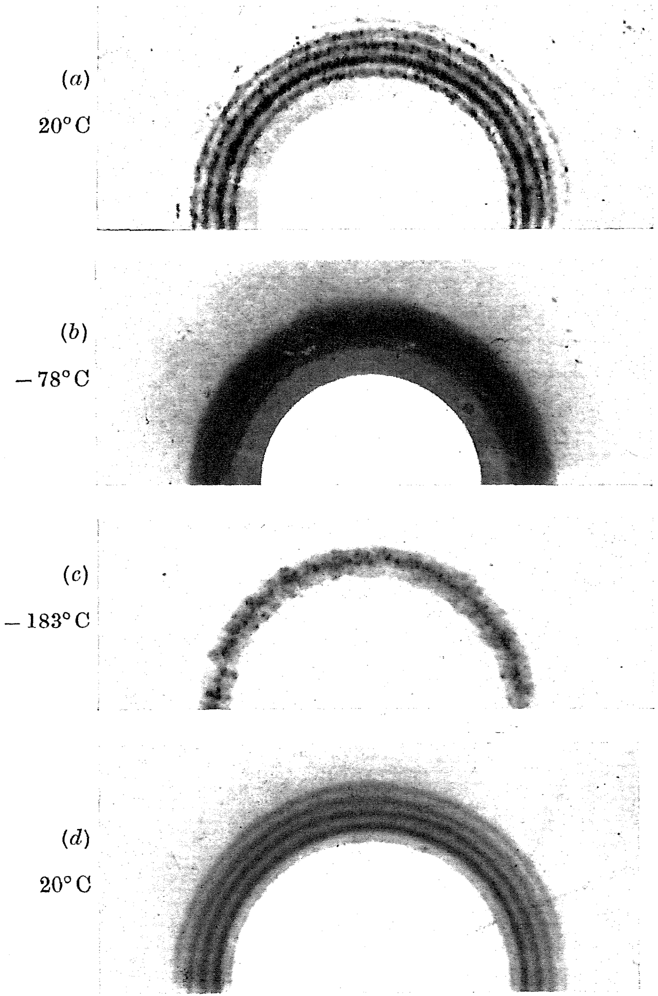
For the low-temperature transition, any suggestions must remain very tentative, as the experimental evidence is so incomplete. However, a picture analogous to that for the high-temperature transition appears to be plausible. The breaking of the *c*-bonds increases the strength of the *a*-bonds. These will not allow movement out of the (001) plane, which would increase their length, but instead give local displacements *in* the plane, regions where the shorter spacing predominates alternating with others where the *c*-bonds are unbroken and the *a*-bonds not shortened. This results in weakening of the *hk*0 reflexions, while the 00*l* are unaffected.

I should like to express my thanks to Dr D. F. Rushman in connexion with whose work this investigation arose, and I am greatly indebted to Sir Lawrence Bragg for permission to use the high-temperature camera in the Crystallographic Laboratory, Cambridge, for a part of the work.

Finally I wish to thank Mr van Moll and the Directors of Philips Lamps Ltd., in whose research laboratory this work was done, for permission to publish this paper.

REFERENCES

Bantle, W. 1942 *Helv. Phys. Acta*, **15**, 373.
Barth, T. 1925 *Norsk. Geol. Tidsskr.* 8, 201.
Bradley, A. J. 1940 *Proc. Phys. Soc.* **52**, 80.
Dinichert, P. 1944 *Helv. Phys. Acta*, **17**, 389.
Dinichert, P. & Weigle, J. 1944 *C.R. Soc. Phys. Genève*, **61**, 173.
Ehrenfest, P. 1933 *Comm. Leiden Suppl.* **75*b***.
Ehrlich, P. 1939 *Z. Elektrochem.* **45**, 362.
Evans, R. C. 1938 *Crystal chemistry*. Camb. Univ. Press.



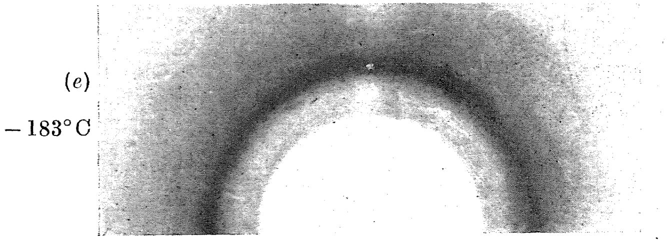


FIGURE 9*a–e*

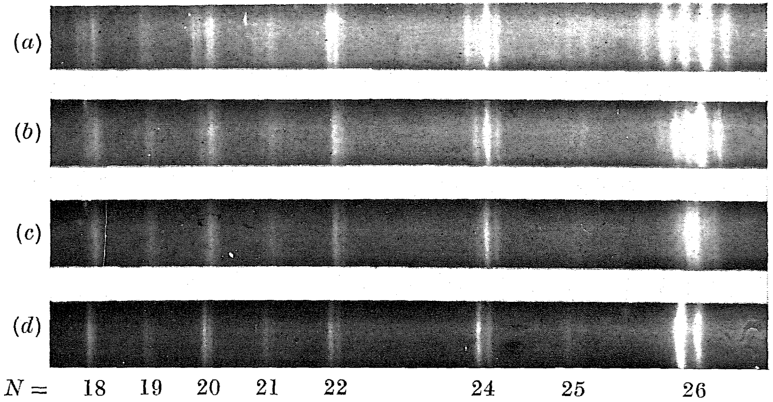


FIGURE 2*a–d*

Huggins, M. L. 1937 *J. Chem. Phys.* **5**, 527.
Jay, A. H. 1933 *Proc. Roy. Soc.* A, **142**, 237.
Klemm, W. 1939 *Z. Elektrochem.* **45**, 583.
Megaw, H. D. 1945 *Nature*, **155**, 484.
Megaw, H. D. 1946*a* *Nature*, **157**, 20.
Megaw, H. D. 1946*b* *Proc. Phys. Soc.* **58**, 133.
Naray-Szabo, S. 1943 *Naturwissenschaften*, **31**, 202.
Pauling, L. 1939 *Nature of the chemical bond.* New York: Cornell Univ. Press.
Pauling, L. & Huggins, M. L. 1934 *Z. Kristallogr.* **87**, 205.
Quervain, M. de 1944 *Helv. Phys. Acta*, **17**, 509.
Rooksby, H. P. 1945 *Nature*, **155**, 484.
Rushman, D. F. & Strivens, M. A. 1946 Faraday Society Discussion on Dielectrics. (In the Press.)
Ubbelohde, A. R. & Woodward, I. 1945*a* *Nature*, **155**, 170.
Ubbelohde, A. R. & Woodward, I. 1945*b* *Nature*, **156**, 20.
Ubbelohde, A. R. & Woodward, I. 1946 *Proc. Roy. Soc.* A, **185**, 448.
Wilson, A. J. C. 1941 *Proc. Phys. Soc.* **53**, 235.
Wul, B. 1945 *Nature*, **156**, 480.
Wul, B. 1946*a* *Nature*, **157**, 808.
Wul, B. 1946*b* *J. Phys. U.S.S.R.* **9**, 95.
Wul, B. M. & Goldman, I. 1945*a* *C.R. Acad. Sci. U.R.S.S.* **46**, 139.
Wul, B. M. & Goldman, I. 1945*b* *C.R. Acad. Sci. U.R.S.S.* **49**, 177.
Wul, B. M. & Vereshchagin, L. F. 1945 *C.R. Acad. Sci. U.R.S.S.* **48**, 634.

DESCRIPTION OF PLATE 5

FIGURE 2. High angle lines on powder photographs of $BaTiO_3$ (Cu $K\alpha$ radiation). (*a*) 20° C, (*b*) 100° C, (*c*) 117° C, (*d*) 200° C.

FIGURE 9. Powder photographs of $BaTiO_3$, to show low temperature changes. (Back reflexion on flat film with Cu $K\alpha$ radiation.) (*a*) Specimen stationary, temperature about 20° C; (*b*) same specimen as (*a*), stationary, temperature about −78° C; (*c*) same specimen as (*a*), stationary, temperature about −183° C; (*d*) specimen rotated, temperature about 20° C (a tungsten wire was embedded in the specimen, and gives a line overlapping with the outermost line of the group $N=26$ due to $BaTiO_3$); (*e*) same specimen as (*d*), rotated, temperature about −183° C (the line due to the tungsten wire is visible, marked *W*).

Visual adaptation in relation to brief conditioning stimuli

By B. H. Crawford, M.Sc.

(*Communicated by T. Smith, F.R.S.—Received* 24 *July* 1946)

(The full text of this paper is printed in *Proc. Roy. Soc.* B, **134**, 283.)

Abstract

Measurements are made of the changes in state of visual adaptation before and after a change in conditioning stimulus. Liminal test stimulus is taken as a measure of visual adaptation.

1. The liminal test stimulus starts to rise sharply about 0·1 sec. before the eye is exposed to the conditioning stimulus. There is also a corresponding, but less marked, rise before the conditioning stimulus is cut off. The remaining sections refer to recovery of dark adaptation after conditioning stimuli of brief duration.

2. The course of recovery of dark adaptation is the same for both positive and negative contrast of the test stimulus against the general background, the numerical values of the contrast being equal.

3. The effect on recovery of dark adaptation of a steady illumination of the field of view is found to be nil until the state of adaptation corresponding to the field illumination is reached, when recovery ceases.

4. The effect of area of test stimulus on the shape of curves of recovery of dark adaptation may be eliminated by the equivalent background brightness transformation: the curves of recovery of equivalent background brightness are coincident for all areas of test stimulus.

5. Recovery of dark adaptation was measured under standard conditions for twenty-six subjects. Personal variation was not reduced by making the equivalent background transformation, indicating that the variation is due to real differences in rate of recovery of the photochemical mechanism of the eye.

6. Recovery curves are given for two subjects for a wide range of intensities of initial conditioning stimulus.

7. A comparison of recovery curves after the two types of conditioning stimulus—a flash covering a large central area of the field of view; a flash covering a small excentric area of the field of view—shows that the equivalent background formula, $\beta = 16E/\theta^2$, already well established for static conditions, applies to this case also. (β is equivalent background brightness, E is illumination at the subject's eye, θ is the angle subtended at the eye between conditioning and test stimuli.)

8. The effect of a succession of conditioning stimuli has been reduced to a simple formula, at least up to a total duration of the succession of 5 sec. Up to 0·3 sec. the eye integrates perfectly the series of conditioning stimuli. Afterwards, integration becomes progressively less perfect according to the relation

$$\epsilon = E\left(\frac{p_0}{p}\right)^{\frac{1}{2}},$$

where ϵ is the intensity of the equivalent single stimulus, E is the total intensity of the multiple stimulus, $p0$ is 0·3, and p is the total duration in seconds of the multiple stimulus.

9. Experiments with conditioning stimuli of different colours show that, within the range of these experiments, it is the scotopic brightness of stimulus which is related to subsequent recovery of dark adaptation, not the photopic brightness.

10. Experiments in the laboratory with natural test objects (lantern slides of various scenes) show that, when an object is on the limit of visibility, the average background brightness in the immediate neighbourhood of the object may be taken as an equivalent background brightness for the calculation of recovery times: a detailed consideration of the pattern of the object itself is unnecessary.

11. The effect of the area of the conditioning stimulus, when this is neither very large nor very small, is investigated. It is found that, under given conditions of observation, each subject has a characteristic critical angle by which he can deflect his fixation from the test object and still perform the visual task required of him. In the experiments here described the visual task was to find the test object and re-align upon it some form of gun sight.

Effect of crossed electric and magnetic fields on the helium spectrum*

By J. S. Foster, F.R.S., *McGill University*, and
E. R. Pounder, Ph.D., *Royal Canadian Air Force*

(*Received* 26 *February* 1946)

[Plate 6]

The authors describe a double canal-ray source, and discuss the general features of greatly improved analyses obtained through its use.

A small canal-ray tube is placed along the axis of a hollow core of the magnet, the pole tip itself serving as the cathode with one small central perforation. A similar tube is symmetrically placed in the other core so that the double canal-ray tube provides opposing beams in the region between the pole faces. Here small field plates are set up to establish strong electric fields perpendicular to the canal rays. Light is taken out at right angles to both fields and analyzed by a six-prism glass spectrograph.

Displacements and qualitative intensities are tabulated mainly for components of three line groups built round and including $\lambda\lambda$ 4922, 4472 and 4026 respectively. $H = 25{,}800$ oersteds while $E = 10$–131 kV/cm.

Resolved components remain sharp at all field strengths. A few components correspond to $\Delta m = \pm 2$; but there is no evidence for $\Delta m = \pm 3$. Intensities of many components are critically dependent upon electric field strength, and in at least one notable case (2^3P–5^3G) displacements are similarly dependent. The latter is attributed to repulsion between levels of different m values which nevertheless are forced to cross over each other.

Introduction

Effects due to uniform external electric fields are best seen in the parhelium spectrum. The observed patterns (Foster 1927) are constant within each spectral series. As interpreted by quantum mechanics, they show that pairs of levels $(n, l, \pm m)$ are coincident and that the selection rule $\Delta l = \pm 1$ no longer applies, since all electric combination lines appear at suitable fields, while $\Delta m = 0, \pm 1$ still holds.

Displacements and intensities are in rather good agreement with the theory, but some intensities show variations, probably due to unequal populations in initial states, and absorption within the source, which are not theoretically considered.

In the present work, a helium source is subjected to uniform electric and magnetic fields crossed at right angles, while the light is taken to the spectrograph in a direction perpendicular to each field. The observations show that the selection rule $\Delta m = 0, \pm 1$ is now extended, since transitions occur for which $|\Delta m| = 2$. It is equally clear—from the absence of certain components—that $|\Delta m| \neq 3$ in any known case.

Professor Bohr (1918) interpreted certain outer components of H_β in a supposedly pure Zeeman analysis by Paschen and Back, as actually arising from a component of electric field directed at right angles to the applied magnetic field. This has been experimentally confirmed. It was not, however, until the introduction of quantum mechanics that uniformly sharp components were theoretically expected to appear in crossed fields of any strength. On this basis the theory of the present problem has been worked out at Professor Heisenberg's Institute.*

---

* The research described in this paper was completed in 1937, prior to service in the Royal Canadian Air Force by one of the authors.

On the experimental side, the problem is to establish and maintain steady electric and magnetic fields in a source of sufficient strength to permit analysis by an optical system of high resolution. Provided the design of the source is such as to give a suitable magnetic field, there is no further difficulty in keeping the field steady. On the other hand, electric fields in all strong sources depend to some extent upon space charges which may vary during the exposure. A major part of the experimental problem is therefore similar to that encountered in the application of electric fields alone.

From earlier experiments with crossed fields (Foster 1931) it is already known that in the sharp and principal series, where the effect of the electric field is small, there is a nearly normal Zeeman separation. In the *P-P* series, however, the observed separations were much less than normal while the diffuse series show more and the *P-F* series much more than normal Zeeman splitting in low electric fields. These wider structures are obviously complex and only partially resolved. They are asymmetric in intensities and show large changes in relative intensities with increasing electric field.

Steubing & Redpenning (1935) have published a more extended analysis of the helium spectrum in crossed fields. Owing to the presence of unsteady electric fields, these authors have found diffuse components at high electric fields where many lines fade out. Since these components are still sharp in fields higher than those established by Steubing & Redpenning, no confirmation of this work can be claimed.

THE LIGHT SOURCE AND EXPERIMENTAL PROCEDURE

Obviously, from what has been said, the design of a suitable discharge tube is an important matter. The tube used in the present experiments ran along the axis of the magnetic field. Briefly, it consisted of two canal-ray tubes opposing each other with the result that the two streams of canal rays (after passing through the pole pieces) ran in opposite directions through the central region between the poles. Here separate field plates were used to establish an electric field at right angles to the streams. Light was conveniently taken out through a side tube at right angles to both electric and magnetic fields. Under operating conditions, each canal-ray beam was actually projected for a short distance into the other discharge tube. The oscillating positive ions thus aided ionization and the production of light.

Looking now at the central portion of the tube in more detail (figure 1), the two pole tips are soldered to an intermediate section of brass so shaped as to form with the pole tips a cylindrical unit in which the pole faces have a fixed separation of 7 mm. Steel shoulders at either end of this part of the tube prevent any motion which otherwise might arise from unbalanced forces on the pole tips. A 0·06 in. hole drilled along the axis of this central structure clears the canal-ray beams which are limited to 0·04 in. diameter by perforated aluminium cathode disks. Owing to the small hole through the steel, as required on this plan, the magnetic field realized in the experiments is about 80 % of that obtained with solid poles. The aluminium disks not only limit the diameter of the beams—and thereby prevent bombardment of the canal wall and electric field plates—but also minimize sputtering. Resting on these disks, and closely fitted into the steel sleeve extensions from the pole tips, are the glass tubes which support the anodes.

---

* Private communication.

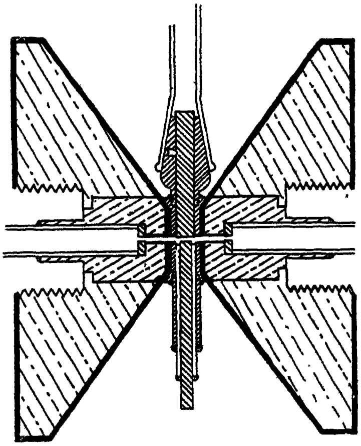


FIGURE 1. Double canal-ray tube (anodes not shown).

Returning now to the centre of the tube, a 0·200 in. hole is drilled in the brass cylinder along a diameter. This hole is threaded for a short distance at each end, and is designed to receive the electric field plates and the gas connexion. At one end the $\frac{1}{8}$ in. positive field plate is fitted with its insulating glass sleeve, and the whole is supported by an outer brass pipe threaded into the main tube. Into the other end the grounded plate is fitted, after having been channelled to permit flow of helium. This plate is also supported by an outer pipe threaded into the main tube. In this case, however, the rather heavy outer pipe is tapered to receive a ground glass connexion to the vacuum system. The grub screw which fixes the position of the field plate is sealed by the tapered joint.

Finally, a central 0·080 in. hole is bored perpendicular to the one which mounts the field plates and this is suitably reamed to form a conical hole through which the light passes out to the spectrograph. Thus the observations are made in a direction perpendicular to both applied fields. The side tube is sealed off by a glass plate fitted over a flattened portion of the brass cylinder. Seals were made with suitable grades of wax and cement.

On the way to a six-prism glass prism spectrograph the light passed through a Wollaston prism. A quarter-wave plate was inserted into the beam with vertical electric vector in order to reduce losses at the prism faces.

With the magnetic axis vertical, it was convenient to lift the upper core of the magnet with block and tackle and insert the tube. The anodes were connected in parallel to a Kipp and Zonen rectifier ('proton apparatus') with resistance control in the primary. The electric field plates were attached to a second rectifier with Variac control.

The tube could be sufficiently outgassed in a few hours. It was then possible to keep a steady voltage of 10,000 to 12,000 on the anodes, while the electric field plates—with 0·75 mm. separation—drew no measurable current at 10,000 V. Since the tube could be operated at 0·7 mm. of purified helium, the double canal-ray feature provided light of sufficient intensity even though the current did not exceed 1 mA.

Exposures varied from 1 to 3 hr. The magnetic field was 25,800 oersteds, while the electric field varied in different experiments with a maximum of 131,000 V/cm. The tubes were commonly good for a few runs but eventually failed owing to disintegration of the glass tubes near the cathode surfaces. Other occasional failures arose from breakdown of electric field plate insulation, or melting of wax seals.

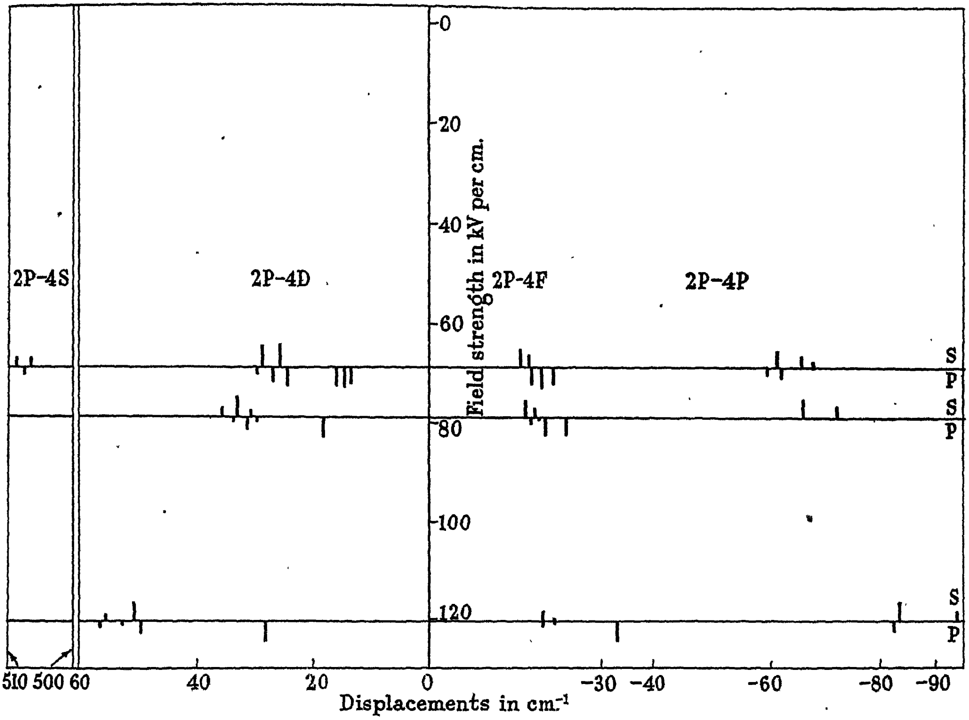
DETERMINATION OF FIELD STRENGTHS

Owing to the method of fixing the separation of the pole pieces, there is no reason to expect the magnetic field to vary except through the current in the coils. This was kept constant, with the result that the normal Zeeman effects showed a field of 25,800 oersteds with a possible error of 2%.

Electric fields cannot be so directly determined, since the experiments show that space charges appreciably modify the field one might expect from the voltage across plates of known separation. Since, however, in most of our experiments the displacements due to the electric field are very large compared with those due to the magnetic field, one may consider the centre of gravity of the simpler complex structures as having the theoretical displacement arising from electric field alone. This assumption leads to rather small variations in calculated field when it is applied to a number of complex structures within the same group of lines. The mean value so obtained is therefore accepted, and is believed to have no greater possible error than that of the magnetic field.

RESULTS

What appear to be the more important measurements obtained in the experiments are recorded in tables 1, 2 and 3. They represent effects in three line groups, 2P–4SDFP (parhelium) (λ4922), 2^3P–4^3SPDF (orthohelium) (λ4471) and 2^3P–5^3SPDFG (orthohelium) (λ4026) respectively. The intensities, as indicated by the length of the lines in figure 2, are very rough estimates. Selected plates are reproduced on figure 3, plate 6. It should be noted that these reproduced



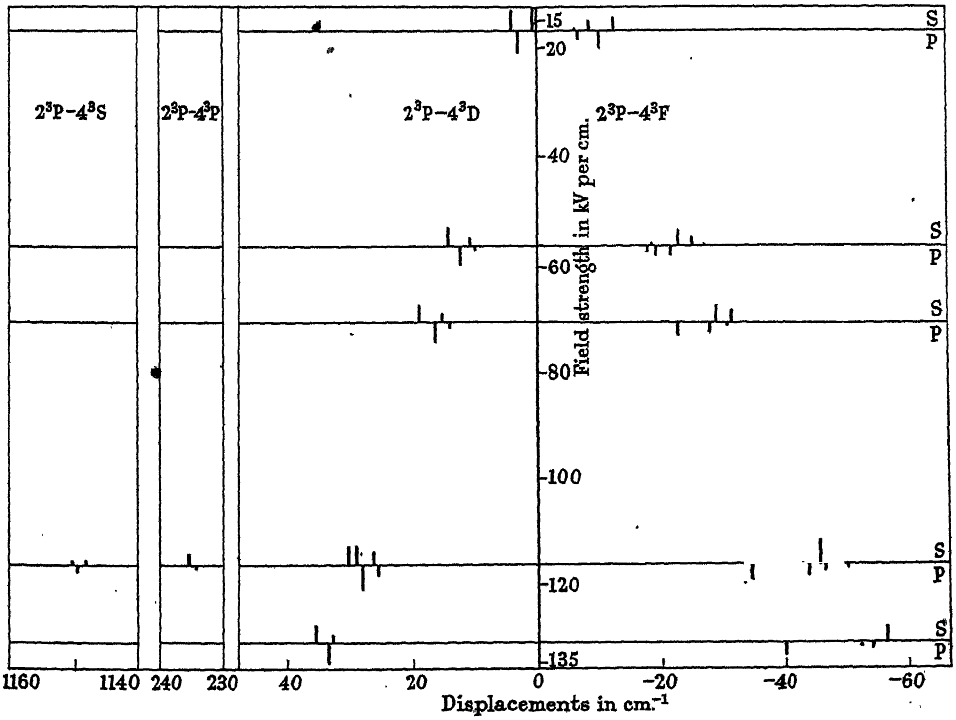


FIGURE 2. Upper: displacement-field strength plot for λ4922 group.
Lower: same for λ4472 group.

photographs cannot be compared exactly since the dispersion on various plates ranged between 14 and 27 cm.$^{-1}$/mm. Further, the enlargement factor is not exactly constant for all prints.

In the tables and reproductions of the photographs the two polarizations are referred to somewhat arbitrarily as perpendicular and parallel. The '*s*' polarization designates that polarization which is perpendicular to the magnetic field and parallel to the electric field. The '*p*' polarization is the reverse of this.

TABLE 1. DISPLACEMENT OF COMPONENTS OF 2P–4SDFP (PARHELIUM) GROUP FROM NORMAL D LINE IN CM.$^{-1}$ AT SEVERAL ELECTRIC FIELD STRENGTHS

magnetic field strength $H = 25{,}800$ oersteds
normal D line is 4921·926 I.A. = 20,311·595 cm.$^{-1}$
dispersion at 4922 A = 23·8 cm.$^{-1}$/mm.

		field strength in kV/cm.		
		68·7	78·8	120·2
polarization	line	displacements in cm.$^{-1}$		
	2P–4S	+509·3	—	—
		506·9	—	—
	2P–4D	28·4	35·9	55·6
			33·0	50·6
		+25·6	30·7	—
	2P–4F	−15·9	16·8	19·9
		17·4	18·7	21·9
	2P–4P	61·2	66·0	83·5
		65·5	72·0	94·0
		−67·5	—	—
p	2P–4S	+507·9	—	—
	2P–4D	29·6	33·9	56·6
		26·9	31·6	52·4
		24·4	29·7	49·5
		15·7	18·6	28·1
		14·5		
		+13·6		
	2P–4F	—	17·7	—
		−18·0	19·0	19·9
		19·5	20·3	21·9
		21·5	23·7	32·8
	2P–4P	59·5	—	82·5
		61·8	—	

DISCUSSION OF RESULTS

The electric field brings out the well-known new electric combination lines which join with lines of the normal spectrum to form groups with a marked overall symmetry. This symmetry arises from interaction (repulsion) between sub-levels with the same *n*, *m* values in electric field alone. In crossed electric and magnetic fields there is interaction also between levels of different m values. These interactions produce separations between certain levels which are quite comparable to Zeeman separations. On this basis, normal Zeeman separations no longer are uniformly maintained. Even in high electric fields the somewhat complex structures continue to show rapid variations of relative displacements with electric field. The resulting structures in most cases are beyond complete resolution in these experiments. Parenthetically, it should be noted that the weakest member of the orthohelium triplet failed to appear.

TABLE 2. DISPLACEMENT OF COMPONENTS OF 2^3P–4^3SPDF (ORTHOHELIUM) GROUP FROM NORMAL d LINE IN CM.$^{-1}$ AT SEVERAL ELECTRIC FIELD STRENGTHS

magnetic field strength $H = 25{,}800$ oersteds
normal D line is 4471·48 I.A. = 22,357·707 cm.$^{-1}$.
dispersion at 4471 A = 20·3 cm.$^{-1}$/mm.

		field strength in kV/cm.				
		17	56·2	70·4	116·1	131
polarization	line	displacements in cm.$^{-1}$				
s	2P–4S	—	—	—	1150·6	—
		—	—	—	1148·1	—
	2P–4P	—	—	—	235·5	—
	2P–4D	+3·7	14·1	18·7	30·0	35·6
		+0·4	10·7	15·1	28·8 26·3	32·8
	2P–4F	**−6·1**	**18·4**	—	—	—
		−8·4	22·6	28·8	45·6	—
		−12·2	24·9	31·2	49·2	56·6
p	2P–4S	—	—	—	1149·5	—
	2P–4P	—	—	—	234·4	—
	2P–4D	+2·9	+12·2	+16·3	+28·0	—
			10·0	13·9	25·6	33·6
	2P–4F	**−6·7**	**17·8**	22·6	−33·3 34·5	40·0
		−10·1	**19·0** 21·6	27·6	44·0 46·4	
			—	30·4	50·1	54·3

PROOF THAT $|\Delta m| = 0, 1, 2$

Structures of the simplest type offer the best opportunity for a clear interpretation of some of the results. These are the components arising from transitions in which the initial states have m values ± 2. For known reasons which cannot be briefly stated, some of these stand well apart from their neighbours under electric fields alone, and this feature is retained in the present photographs. Here they may be expected to provide two components only ($m = 2 \rightarrow m = 1$; $m = -2 \rightarrow m = -1$) on the basis of the usual selection rule $\Delta m = 0, 1$.

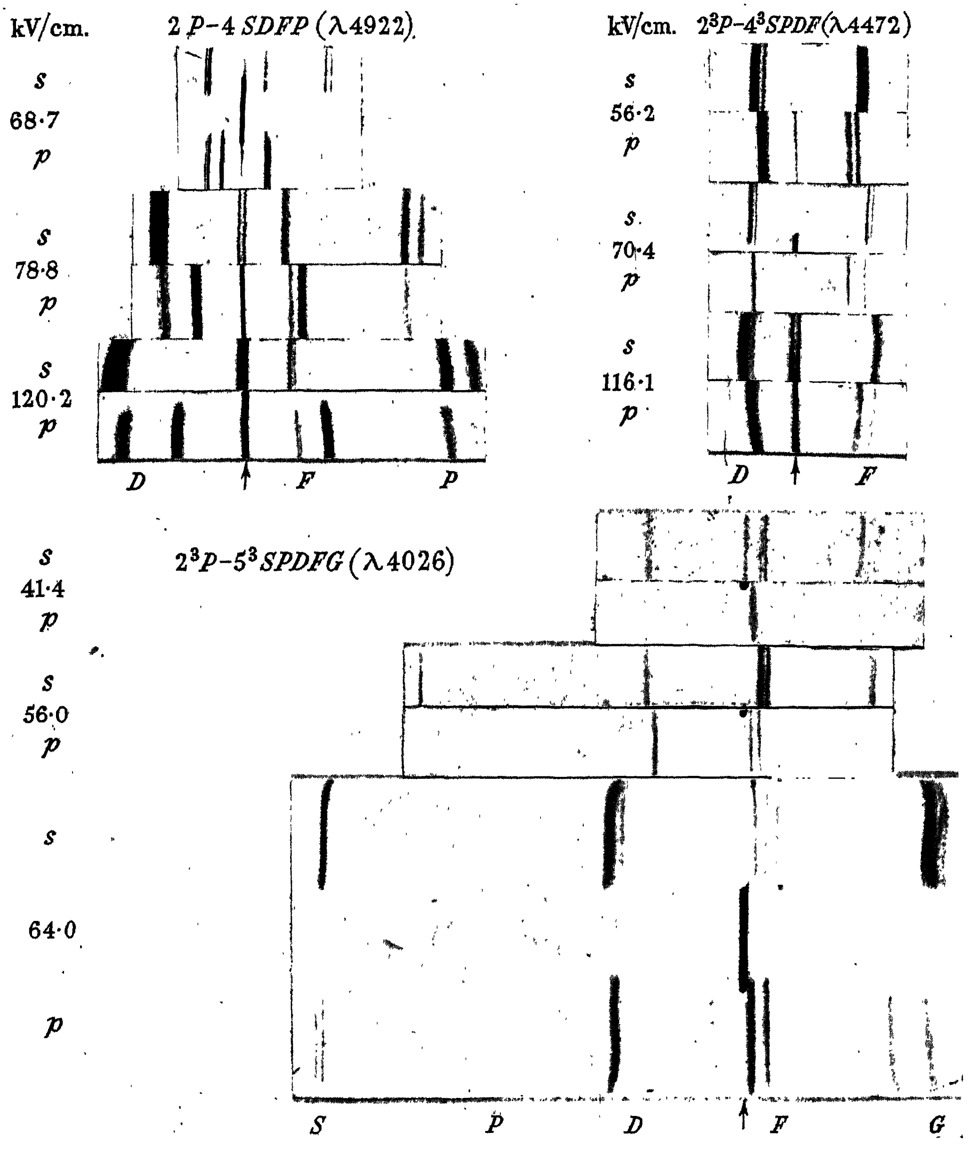
Actually, only three examples of such components with suitably clear separations and intensities are found. The corresponding observations are set in heavy type in the tables. Each structure contains at least *three* members of different wavelengths. Except in the first example, the polarization is not the same for the three components. These observations mean that in all cases where clear examination is possible there are transitions in which m changes from ± 2 to 0. It may be concluded that in crossed fields

$$|\Delta m| = 0, 1, 2.$$

TABLE 3. DISPLACEMENT OF COMPONENTS OF 2^3P–5^3SPDFG (λ4026) (ORTHO-HELIUM) GROUP FROM NORMAL D LINE IN CM.$^{-1}$ AT SEVERAL ELECTRIC FIELD STRENGTHS

magnetic field strength $H = 25{,}800$ oersteds
normal D line is 4026·190 I.A. = 24,830·38 cm.$^{-1}$
dispersion at 4026 A = 15·6 cm.$^{-1}$/mm.

		field strength in kV/cm.			
		41·4	56·0	64·0	81
polarization	line	displacements in cm.$^{-1}$			
s	2P–5S	—	—	+576·66	—
		—	—	574·24	—
	2P–5P	—	120·7	139·8	—
		—	—	—	133·8
		—	—	—	121·8
		—	—	—	117·2
	2P–5D	26·2	35·4	40·1	50·9
			33·6	38·6	
		+24·2	30·9	+35·8	46·8
	2P–5F	—	—	**−1·79**	—
		−2·31	2·98	**2·70**	—
		4·60	7·47	7·50	8·91
		5·94	10·81	10·52	12·10
	2P–5G	31·4	—	52·3	67·5
		34·0	—	55·0	71·9
				56·1	
		—	—	−58·4	—
p	2P–5S	—	—	+575·58	—
	2P–5P	—	122·0	125·3	—
		—	119·8	123·0	—
		—	—	—	119·4
	2P–5D	+22·8	32·4	+37·2	48·8
	2P–5F	−2·00	1·89	**−1·90**	—
		3·44	3·52	3·38	—
		—	5·52	5·93	—
		—	6·72	7·36	—
	2P–5G	—	39·6	43·6	—
		—	48·6	52·3	—
				−54·3	—
		—	—	—	−78·0



$2P-4SDFP$ ($\lambda 4922$)

2^3P-4^3SPDF ($\lambda 4472$)

2^3P-5^3SPDFG ($\lambda 4026$)

FIGURE 3. $H = 25{,}800$ oersteds in all cases. E values as indicated. In each set the arrow indicates the position of the undisplaced line.

It is, however, equally clear that there are no observed components for which $|\Delta m| = 3$. Obviously these could possibly arise only from initial levels 4F, 4^3F, 5^3G without electrical displacement, and 5^3FG with relatively small displacements. There is no question of adequate resolution of the components, should they appear. The photographs show that the transitions $|\Delta m| = 3$ do not take place and incidentally that the closely associated transitions ($m = 3 \rightarrow m = 1; m = -3 \rightarrow m = -1$) where $|\Delta m| = 2$ are also missing.

We shall next consider more carefully a second new feature in crossed fields, viz. a continuous interaction between terms, notably within the same close structure, which leads to many striking changes in intensities and displacements with increasing electric fields. An outstanding example is found in the '*s*' polarization of 2^3P–5^3G. While the plates show that the analysis obviously changes greatly with large variations of the electric field, the observations near 64 kV/cm. are of particular interest.

In this photograph (the bottom one on figure 3, plate 6) it would at first seem that the electric field had varied abruptly during the exposure and that a pair of spurious components were so produced. This is certainly not the case, neither may the effect be attributed to final states, since similar features are not found for other lines of comparable displacement and intensity. Indeed some well-displaced components are so sharp that on second consideration it seems certain that the source was exceptionally steady with only an important small variation in electric field from point to point along the axis of the tube.

From the above one may conclude that this particular effect is mainly in the 5^3G states and at the same time observe that the components from initial sub-levels $m = \pm 2$ are well separated in '*p*' polarization and therefore are not of prime importance in the discussion. This leaves mainly the interaction between 5^3G sub-levels with $m = 0, \pm 1$. Indeed, the facts strongly suggest that the $m = 0$ level moves from an inside position across the $m = -1$ level, to the outside as a result of the very small range of electric fields represented.

Assuming the above qualitative explanation to be correct, it is clear that such effects will be detected on comparatively few plates and only under the most exact controls. This case emphasizes the ease with which components may be blurred in crossed fields. Evidently the demands upon the experimenter are here somewhat greater than is encountered in work with either field alone.

The remainder of the fine details are left to the reader for examination. It may be remarked that the present method is suitable for extensions. It would help to have higher magnetic fields, and slightly inclined electric field plates, to give some variations which would assist one in making positive identification of many more components.

REFERENCES

Bohr, N. 1918 *K. Danske vidensk. Selsk. Skr.* ser. 8, **1**, 1.
Foster, J. S. 1927 *Proc. Roy. Soc.* A, **117**, 137–163.
Foster, J. S. 1931 *Proc. Roy. Soc.* A, **131**, 133–146.
Steubing & Redpenning 1935 *Ann. Phys., Lpz.*, **24**, 161.

The surprising amount of energy which can be collected from gases after the electric discharge has passed

By Lord Rayleigh, F.R.S.

(*Received* 14 *June* 1946)

It was found in earlier work (Rayleigh 1940 *b*) that a piece of gold sheet, say 1 cm. square, hung up in a stream of glowing active nitrogen became red hot, and collected an amount of energy which was very surprising. It was necessary to assume that every molecule of nitrogen in the stream imparted 10 eV to the gold plate.

In this form of experiment it was impossible to avoid large and undetermined losses of energy; thus the energy actually given up by each molecule was probably much larger still.

A new form of experiment is now described in which this loss is avoided. A platinum strip is kept hot by periodic discharges, and by making the experiment (1) with the platinum exposed to the gas and (2) with the platinum protected by a thin glass sheath it is possible to determine what part of the total energy is to be attributed to catalytic action of the discharge products. This amount of energy is found to be very great. Reckoning in electron-volts per molecule of gas present it increased rapidly as the gas pressure was lowered and at the lowest pressure used it rose as high as 223 eV/mol. Results of the same order were obtained with other gases, so it is not clear that the glow of active nitrogen is the essential condition. This great liberation of energy much exceeds what can be explained by dissociation of the molecules and single ionization of every atom which results, which would only afford 36 V.

In a former paper (Rayleigh 1940 *b*) I showed that metal plates, notably gold, silver, copper and platinum, hung up in a bulb containing glowing active nitrogen, but right away from the discharge which maintained the activity of the gas, would become red-hot, and that plates 1 cm. square of gold, silver and copper could even be made to melt, and run down.* This was made the basis of a minimum estimate of the energy contained in active gas. A stream of it was maintained by an air pump, so that the volume passed could be compared with the energy liberated on the metal. The work was of an exploratory nature, and the methods used were crude, but they indicated that the energy which was collected from the gas was so great that *every molecule in it* must be excited to the extent of at least 10 eV.

It was shown further that this energy was of the order of a thousand times the energy radiated under the most favourable conditions as visual nitrogen after-glow.

As regards the energy collected by the metal, the form of experiment then used has obvious weak points. The gas has to travel a considerable distance before it gets to the metal where its energy is given up, and in the course of doing so it is exposed to considerable areas of glass wall, with which its molecules may make repeated collisions. There is, moreover, a time lag while it travels considerable distances, and during this period decay of activity may occur. [If the reader is inclined to complain that the word activity is here used somewhat vaguely it may be defined simply as the capacity to impart energy to metals like gold and platinum initially cold.]

Another weak point was that the attempt to use the visual incandescence of the metal as an indicator of temperature was crude. It became obvious that a method of electrical resistance thermometry would be much better, and, indeed, this would have been used had it not been that the uncertainties of the time (1939–40) made it desirable to achieve what progress I could at once, without taking time to change the method.

---

* *Note added in proof.* I have since melted a fine platinum wire in the same way.

In planning further experiments, it was desired to bring the metal catalyst as near as possible to the discharge where the gas was made active, so that it would have the minimum opportunity of losing its activity. At the same time it was necessary to make allowance for the direct-heating effect of the discharge on the catalytic metal, which cannot be altogether avoided without making the gas travel considerable distances—and this was inadmissible.

To produce the electrodeless discharges, the same oil condenser was used as in my previous work. It was charged by an induction coil with a slow mercury break. This break was generally of a design due to the late Sir James Mackenzie Davidson. An ebonite wheel with a conducting segment is mounted on an axle which is inclined at 45° to the horizontal. This wheel is immersed in mercury and makes and breaks contact once a revolution. The mercury is covered with a layer of alcohol.

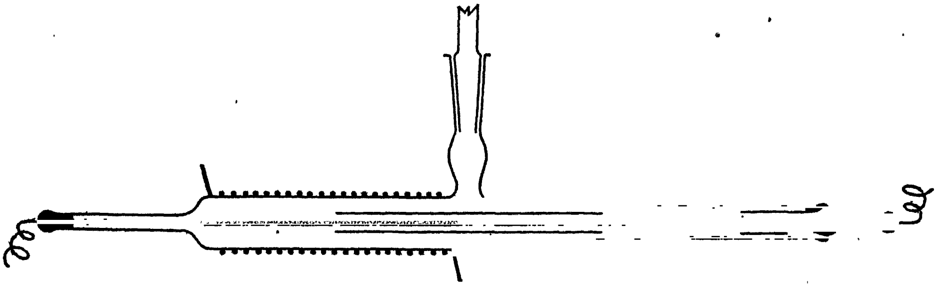


FIGURE 1. Reduction $\frac{2}{5}$.

The discharge vessel is as shown in figure 1. Along the axis there is stretched a platinum strip, over which a glass sheath can be slipped. The platinum can be heated by a current from a battery, and the tube is wound with a wire coil, which is used to excite the electrodeless discharge. When the discharge passes, the platinum strip gets heated, and its resistance increases. This is attributed partly to the direct-heating action of the discharge, and partly to the action of the platinum strip in catalyzing the discharge products (ions and dissociated atoms?).

During an experiment, the platinum strip is kept at an arbitrary but constant resistance of R ohms. For this purpose it is made one arm of a Wheatstone's bridge, and the current adjusted to heat the strip until the given resistance is attained. The current (C_1 amp.) is measured. Then the glass sheath is slipped into position, so as to screen the wire from dissociation products. A larger current (C_2 amp.) must now be passed to restore the resistance to its standard value of R ohms. The energy given up to the platinum wire by the dissociation products is

$$(C_2^2 - C_1^2)\,Rt\ \text{W-sec.},$$

when t is the time interval in seconds from one discharge to the next.

This energy is derived from, and therefore is contained in, the gas volume v at pressure p cm. of mercury.

Thus the energy per unit volume reckoned as at atmospheric pressure is

$$\frac{(C_2^2-C_1^2)\,Rt\times 76}{vp}\ \text{W-sec./c.c.}$$

In a typical experiment, when the conditions were adjusted to get the best effect at the chosen pressure, $C_1 = 0{\cdot}78$ amp., $C_2 = 1{\cdot}73$ amp., $R = 0{\cdot}470$ ohm., $t = 0{\cdot}25$ sec., $v = 19$ c.c., $p = 2{\cdot}6\times 10^{-3}$ cm. of mercury. In this case the energy is 432 W-sec./c.c. as at N.T.P., or 98·5 eV/mol. of nitrogen present.

All the molecules in the vessel play their part, and must travel from the point in the discharge vessel where they receive energy to a point in the central strip where they give it up. They will not in general take the shortest path, that is, the path along the radius of the vessel. It is not worth while to enter upon elaborate calculations of the most probable length of path. I shall take it as 1 cm., about the same as the radius of the vessel.

If the molecule is to receive energy and give it up more than once during a discharge, it would in general have to travel some 2 cm. while the discharge is actually passing. The duration of discharge (several oscillations) was estimated by photographing a slit backed by its light, in a revolving mirror, and found to be $3{\cdot}76\times 10^{-6}$ sec. This requires the molecule considered to have a velocity of $\frac{2}{3{\cdot}76}\times 10^6$ cm./sec. or say 5×10^5 cm./sec. Now the molecular velocity of nitrogen is 5×10^4 cm./sec. at 273° K. To raise it 10 times, the absolute temperature must be increased 100 times, i.e. to the temperature of the hottest stars. It seems clear, therefore, that a molecule could not travel fast enough to give up energy to the platinum strip more than once during the short duration of a discharge. If not, it must carry 98·5 eV at one time, and it is not easy to see how according to current ideas it could do this. The energy of dissociation of nitrogen is 7·4 (?) V, and this, together with the ionization of both the atoms, would only afford $7{\cdot}4 + 2\times 14{\cdot}5$ V or 36 V. Considering that the discharge gives a band spectrum and not a line spectrum, the idea of complete dissociation and ionization is somewhat fantastic; but even if we make this assumption the difficulty remains.

There appears to be no advantage in increasing the number of turns to the maximum, and close winding was inconvenient, because the coil showed a tendency to spark over. A coil having 26·5 turns was used in further experiments, as being about the useful maximum.

The pressure was then varied, leaving the other conditions unaltered:

pressure (cm.)	W-sec./c.c.	eV/mol. present
26×10^{-3}	5·32	2·16
$12{\cdot}4\times 10^{-3}$	19·1	4·37
$5{\cdot}2\times 10^{-3}$	92·3	21·1
$2{\cdot}6\times 10^{-3}$	309	70·6
$1{\cdot}5\times 10^{-3}$	643	147
$1{\cdot}1\times 10^{-3}$	1020	233

It appears, therefore, that the energy per molecule increases very greatly as the pressure is diminished.

It was of interest to see whether these very large values of the energy per molecule are peculiar to nitrogen, or whether they would be found in other gases also. A few results only are available at present:

gas	F, pressure (cm.)	W-sec./c.c.	eV/mol.
oxygen	$2\cdot6\times10^{-3}$	432	98·5
hydrogen	$4\cdot3\times10^{-3}$	470	107
nitrogen	$2\cdot6\times10^{-3}$	309	70·6

These are of the same order of magnitude as the results for nitrogen, one of which is repeated for convenience of reference.

CONCLUSION

If a platinum surface is exposed to nitrogen through which a discharge has passed, the energy which it collects from the gas is greatly in excess of what would be expected if the nitrogen were completely dissociated into atoms, and every resulting atom was singly ionized. It is difficult to see how the result can be explained by current conceptions. Similar results have been obtained with oxygen and hydrogen.

REFERENCES

Rayleigh, 1940*a* *Proc. Roy. Soc.* A, **176**, 1.
Rayleigh, 1940*b* *Proc. Roy. Soc.* A, **176**, 16.

Fragmentation of shell cases

By N. F. Mott, F.R.S.

(*Received* 14 *December* 1945)

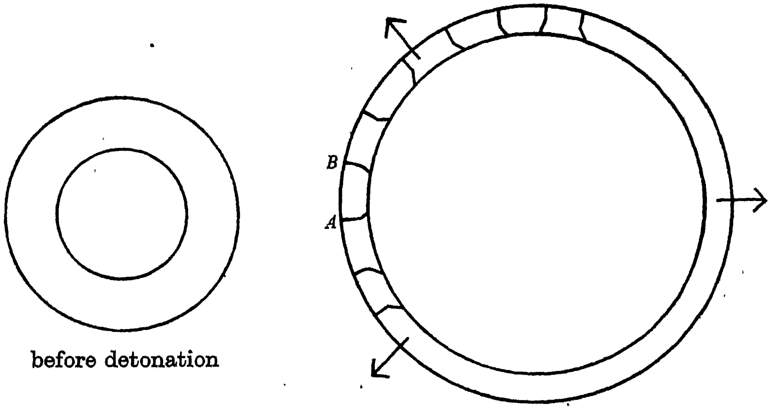
A theory is given of the break-up of the metal case of a cylindrical ring-bomb, in which the lines of fracture perpendicular to the axis of the bomb are predetermined. From the theory an expression is given for the length of the average fragment; this is shown to depend, if certain hypotheses are made, on the radius and velocity of the case at the moment of break-up, and on the mechanical properties of the metal.

1. INTRODUCTION

This paper is the outcome of attempts made by the present author to find a theoretical basis for the prediction of the distribution in weight of the fragments of shell or bomb cases after detonation of the filling. Little attempt will be made here to relate the theory to experiment, because experimental determinations of the distribution of the weights of fragments have not usually been made under the conditions best suited for theoretical interpretation. The purpose of this paper is to present a theory, and to state the kind of experimental work which if carried out would show whether or not it is correct. It is suggested here that the average size of the fragments from a shell case of a ductile metal depends on a property of the metal which is not usually measured, namely, the *scatter* in the values of the reduction in area at which fracture occurs in a tensile test. Determination of this quantity is of interest for a theory of fracture in metals, and measurements of the fragment weight distribution are therefore capable of giving information of fundamental interest about the properties of metals.

When the explosive filling of a shell or bomb detonates, the case is subject initially to an extremely high pressure from the gaseous products of detonation; under this pressure it begins to move rapidly outwards. If the case is made of a ductile material such as steel, very considerable plastic expansion, as much as 50 %, occurs before the case breaks; this is most easily seen by examination of the larger fragments from a shell which include parts of the original inner and outer surfaces; the distance between these surfaces will be less than in the original case. By the time that fracture occurs the velocity with which the case is moving will be nearly equal to the final velocity with which the fragments are projected, and the gas pressure will have dropped to a small fraction of its original value. The fragmentation of a shell case must thus be thought of as the tearing apart of a rapidly expanding tube when the material of the tube reaches the limit of its ductility.

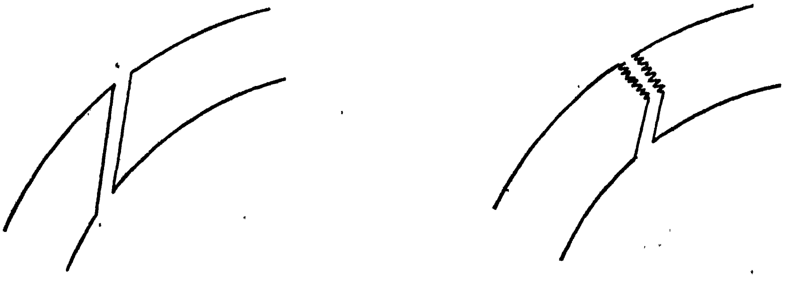
Fracture in forged steel shell cases occurs normally by the formation of cracks parallel to the axis of the cylindrical part of the case, giving the familiar long thin fragments from bursting A.A. shells. A complete theory of fragmentation would have to account for both the width and length of fragments. If, however, the wall ll consists of a number of coaxial circular rings, all with the same inner and ii, stacked one above the other round a thin steel inner lining, the problem ; each ring will break into a number of pieces, and the planes of fracture rallel to the axis of the cylinder. Ring-bombs of this type have been made, l to break up in the way stated here.



before detonation

at moment of fracture

FIGURE 1. Case of shell before detonation and at moment of fracture.



(*a*) (*b*)

FIGURE 2. Types of fracture of bomb casings.

The experimental arrangement to which the present theory will be applied will thus be as follows: The explosive is in the form of a cylindrical tube with a thin metal wall, of length several times the diameter to ensure that a steady detonation wave is set up. The tube is encased in a series of metal rings all of the same dimensions, which fit it closely. The explosive is detonated and the fragments collected. The theoretical problem is to calculate the distribution of lengths (AB in figure 1) which the fragments of the rings will be found to have.

Two main types of fracture have been observed in bomb and shell casings; shear fracture, approximately at 45° to the circumference of the case, as in figure 2*a*, and a combination of fibrous fracture (of the same type as at the bottom of the cup and cone fracture in a tensile test) and shear fracture as in figure 2*b*. In § 3 consideration is given to the mechanism of fracture; for the purpose of the mathematical theory to be developed here, it need only be assumed that fracture, once takes place so rapidly that it can be considered instantaneous compared rate of strain of the metal.

2. THEORY OF FRAGMENT LENGTH

The tensile strength of a brittle material such as glass shows a very co[illegible] scatter, of the order ±20 %, from specimen to specimen. This is becaus[illegible] begins at one of a number of weak spots on the surface (the Griffith cr[illegible] strength is believed to depend on the depth of the deepest crack, and this[illegible] from one specimen to another. The true stress at fracture of ductile mat[illegible] as steel shows much less variation, as does also the reduction in area at w[illegible]ture occurs. No systematic investigation of the magnitude of the scatter has been made, but for specimens cut from the same bar it can be less than ±1 % for a steel showing a total reduction in area of 50 %. Initiation of fracture in metals is not necessarily a surface phenomenon; in a tensile test it starts in the interior of the specimen and little is known of the nature of the points of weakness responsible. Nevertheless, it is unlikely that even in a homogeneous material there will be no scatter. In any case it will be assumed that there does exist a scatter characteristic of the homogeneous material, and not due to inaccuracies in the apparatus. It will be found that the magnitude of this scatter determines the fragment size.

Denote by s the plastic strain of the specimen, so that if the cross-sectional area is A and the original area A_0,

$$s = \frac{A_0 - A}{A}.$$

Assume that the chance that an unfractured specimen of unit length will fracture when the strain increases from s to $s + ds$ is

$$Ce^{\gamma s}\,ds, \tag{1}$$

where C and γ are constants. The exponential expression is chosen as the simplest form which gives a rapid increase from negligible to large values as s increases. With this assumption, the chance p that the specimen breaks before a strain s is reached is given by

$$\frac{dp}{ds} = (1-p)\,Ce^{\gamma s},$$

whence

$$p = 1 - \exp\left\{-\frac{C}{\gamma}e^{\gamma s}\right\}.$$

The average strain for fracture s_0 is given by

$$s_0 = \int_0^\infty s\frac{dp}{ds}\,ds$$

$$= \frac{1}{\gamma}\left\{\log\left(\frac{\gamma}{C}\right) + \epsilon\right\},$$

where

$$\epsilon = \int_{-\infty}^{\infty} xe^{x}\exp\left(-e^{-x}\right)dx = 0{\cdot}577,$$

and the r.m.s. of the scatter in the strains at fracture is

$$\left\{\int_0^\infty (s-s_0)^2 \frac{dp}{ds}\,ds\right\}^{\frac{1}{2}} \simeq \frac{1\cdot 28}{\gamma}.$$

A value of γ equal to 128 would thus give a r.m.s. scatter of 0·01 in the strain at which fracture occurs.

The argument now is, that if γ were infinite so that the strain at fracture were perfectly definite, the expanding shell case would break in all points at once. With the assumption (1), however, there is in any length of the material a finite probability of fracture which increases rapidly as s approaches the critical value s_0. As soon as fracture takes place at one point, stress is released in the neighbourhood, and the unstressed regions spread with a velocity which can be calculated. This is shown in figure 3; a fracture is supposed to have occurred at A and stress has been released in the regions AB and $A'B'$ which are shaded. Fracture can no longer take place in the shaded regions; on the other hand, in the unshaded regions plastic flow is still going on, s is increasing and according to formula (1) fracture becomes more and more likely. The average size of fragment will be determined by the rate at which the shaded regions in figure 3 spread and so prevent further fracture. Thus this rate must be calculated.

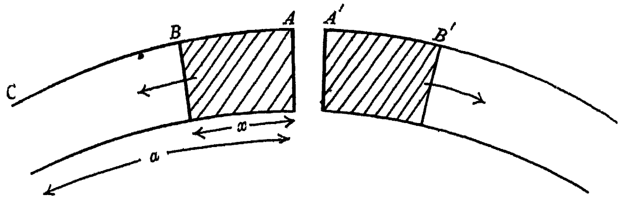


FIGURE 3. Release of strain round a tensile fracture.

Let r be the radius of the case and v its velocity outwards at the moment of break-up, it will be supposed that the thickness of the case is small compared with r, and that the whole process of break-up takes place in a time interval during which there is little change in r or v. Then the rate of strain ds/dt of the material is given by

$$\frac{ds}{dt} = \frac{v}{r}.$$

If in figure 3 an arbitrary point C be taken on the circumference of the case at a distance a from the point of fracture, then relative to C the velocity of the whole stress-free region AB is

$$\frac{v}{r}(a-x),$$

where x is the length AB. If ρ is the density of the material and P_F is its flow stress in tension at the moment of fracture for the rates of strain concerned, then the equation of motion of AB is

$$P_F = -\rho x \frac{d}{dt}\left\{\frac{v}{r}(a-x)\right\},$$

which gives on integrating (v/r being treated as constant)

$$x^2/t = 2rP_F/\rho v. \tag{2}$$

The width x of the unstressed region thus increases as the square root of the time.

This solution treats the metal as inelastic, and gives an infinite value for the initial velocity of the boundary B. A more exact solution* gives a value of the velocity equal initially to that of sound in steel, but except in the opening stages approximating closely to that given by (2).

Making use of (1) and (2), the distribution of fragment lengths can be found. In a ring of diameter l ($= 2\pi r$), suppose that at any moment n fractures have formed. Then the rate of increase of n with increasing strain is given by the equation

$$\frac{dn}{ds} = lfCe^{\gamma s}, \tag{3}$$

where f is the proportion of the ring which is still stressed. It is convenient to introduce a new variable

$$\sigma = \gamma s.$$

Then (3) becomes

$$\frac{dn}{d\sigma} = \frac{lfC}{\gamma} e^{\sigma}, \tag{4}$$

and if a fracture occurs when $\sigma = \sigma_1$, the region round it which is unstressed and thus safe from further fracture (shaded in figure 3) is at any subsequent value of σ, by (2), equal to

$$x_0(\sigma - \sigma_1)^{\frac{1}{2}}, \tag{5}$$

where

$$x_0 = (2P_F/\rho\gamma)^{\frac{1}{2}}\, r/v.$$

The length x_0 is on dimensional grounds obviously proportional to the average fragment length. The distribution of average fragment lengths was found empirically as follows: Using equation (4) it was assumed that the first crack would form when n was unity and thus for a value σ_0 of σ given by

$$\frac{lC}{\gamma} e^{\sigma_0} = 1.$$

A line of length l (say 10 cm.) was ruled on paper to represent the circumference of the cylinder and a mark, representing a fracture, made on it at a point P chosen at random. The next fracture will occur when $n = 2$; before the line is cut at random a second time to represent this happening, a range Δx must be marked off round the first cut to represent the shaded region where the strain has disappeared. By (4) the increase in σ is unity, so the interval marked off on each side is x_0. The second cut must now be placed on any part of the line excluding the shaded region. f is now $(l - 2x_0)/l$; the increase in σ before the third cut is made is by (4) given by

$$\Delta\sigma = 1/fe^{\sigma - \sigma_0} \quad (\sigma - \sigma_0 = 1).$$

---

* E. H. Lee (to be published).

Round the second cut a shaded zone of $x_0(\Delta\sigma)^{\frac{1}{2}}$ must be drawn in, and the first must be widened according to equation (5). The process was repeated until the whole line was covered.

By repeating the whole procedure a number of times and measuring the lengths of the intervals between adjacent cuts on the line, a histogram was drawn giving the numbers of intervals between $0{\cdot}4x_0$ and $0{\cdot}8x_0$, and so on. This is shown in figure 4. The curve drawn through the histogram should give approximately the numbers of fragments with lengths between x and $x+dx$.

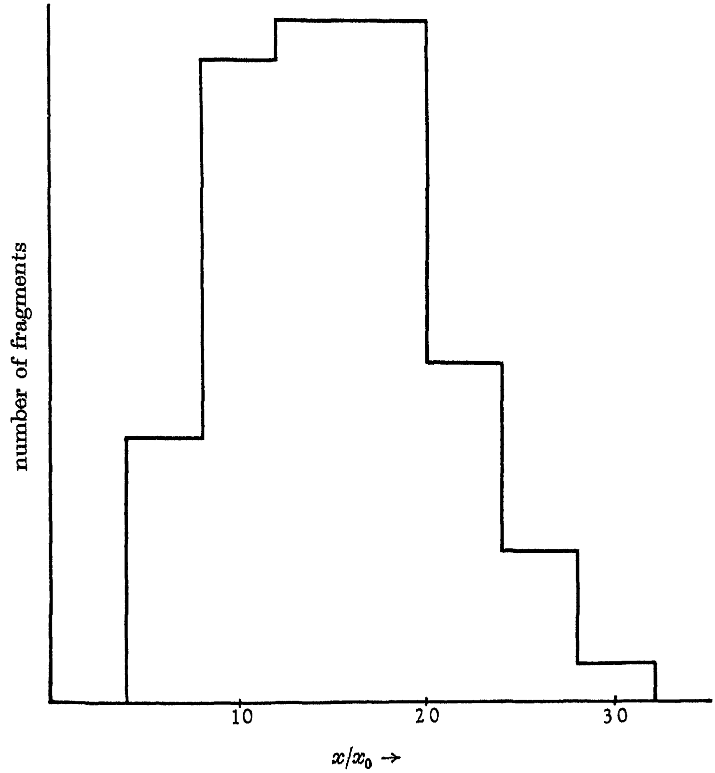


FIGURE 4

The calculations were made with $l/x_0 = 20$; the distribution would not be sensibly different for larger values.

It will be seen that:

(1) The fragments have lengths most of which lie between x_0 and $2x_0$, and that the average length is about $1{\cdot}5x_0$.

(2) x_0 is proportional to r, so that if the linear proportions of the bomb are scaled up by a given factor, the average fragment length is changed by the same factor.

(3) x_0 is inversely proportional to v, the velocity of the case at the moment of break-up.

As a numerical example, consider a bomb of diameter 3 in. so that r will be about 2 in. at the moment of break-up. Suppose that the wall thickness is such that $v = 2000$ ft./sec., and that a steel case with $\rho = 480$ lb./cu.in. has a flow stress in the work-hardened state of 50 tons/sq.in. Then we find

$$x_0 = 1{\cdot}6/\sqrt{\gamma}\,\text{in.}$$

Thus if $\gamma \sim 100$, the average fragment length is about 0·24 in.

3. A THEORETICAL ESTIMATE OF THE CONSTANT γ WHICH DEFINES THE SCATTER IN THE VALUES OF STRAIN AT FRACTURE

In the theory of A. A. Griffith of the strength of brittle solids, it is assumed that fracture starts at the deepest of a number of surface cracks. If c is the depth of a crack, then according to Griffith the stress at which the crack will spread is

$$P = \sqrt{\frac{ET}{\pi c}}, \tag{6}$$

where E is the elastic modulus and T the surface tension. The variation in the depth of the deepest crack determines the variation in the tensile stress from specimen to specimen.

For ductile metals no satisfactory theory exists which accounts for the initiation of fracture at a given tensile stress. It will, however, be assumed that fracture begins, as in brittle materials, at one of a number of 'weak points' distributed throughout the material, and that the relation between the 'size' c of a weak point and the stress required to start a fracture there is of the same type as (6), namely,

$$P = \text{const.}/c^n. \tag{7}$$

Fracture will then occur when, owing to the work-hardening of the material, the maximum principal stress reaches a value great enough to start a crack at the weak point for which c is largest. It will be assumed, moreover that the values of c are distributed about a mean value c_0 according to the Gaussian distribution function; thus the chance per unit volume that there is a weak point with parameter between c and $c+dc$ will be taken to be equal to

$$B\,dc\exp\left[-\frac{(c-c_0)^2}{2\tau^2}\right]. \tag{8}$$

Thus, if a specimen of volume V has not fractured for a smaller stress, the chance that it will fracture for a stress between P and $P+dP$ is

$$BV\exp\left[-\frac{(c-c_0)^2}{2\tau^2}\right]\frac{dc}{dP}dP,$$

where c is given in terms of P by (7).

Interest is centred only in a small range of values of c in the neighbourhood of the largest. Therefore let

$$c = c_R + x,$$

where c_R is defined by $$\int_{c_R}^{\infty} BV\,dc\exp\left[-\frac{(c-c_0)^2}{2\tau^2}\right] = 1. \tag{9}$$

Neglecting terms in x^2, it is found that

$$(c-c_0)^2 = (c_R-c_0)^2+2x(c_R-c_0), \tag{10}$$

and, using this approximation, (9) gives

$$\frac{(c_R-c_0)^2}{2\tau^2} = \log_e\left(\frac{BV\tau^2}{c_R-c_0}\right).$$

Now from (8), integrating over all values of c, it is clear that τB, apart from a numerical factor of order unity, is equal to the total number of weak points per unit volume. This is denoted by N, and N is assumed to be a very large number. Thus the error will be small if

$$\frac{(c_R-c_0)^2}{2\tau^2} = \log_e(NV). \tag{11}$$

If (10) is substituted in (8), then

$$\text{const.}\,dx\exp\left[-\frac{c_R(c_R-c_0)}{\tau^2}\frac{x}{c_R}\right]. \tag{12}$$

A broad distribution of values of c will be assumed, so that τ and c_0 are of the same order of magnitude. Thus from (11) $c_R \gg c_0$. Hence approximately

$$\frac{c_R(c_R-c_0)}{\tau^2} \sim \frac{(c_R-c_0)^2}{\tau^2} = 2\log_e(NV).$$

Thus the chance that the material contains a weak point with parameter x between x and dx is, if x is small

$$\text{const.}\,dx\exp\left[-2\frac{x}{c_R}\log_e(NV)\right].$$

If the stress-strain curve for the material in tension is

$$P = f(s),$$

the constant γ is thus $$\gamma = -2\log(NV)\frac{dc}{c\,dP}\frac{dP}{ds},$$

which gives from (7) $$\gamma = 2\log(NV)\frac{1}{n}\frac{d(\log P)}{ds}.$$

If the stress-strain curve for large strains is of the type

$$P = P_1+P_2\log(1+s),$$

this gives $$\gamma = 2\log(NV)\frac{1}{n}\frac{P_2}{P_F}\frac{1}{1+s_F},$$

where P_F is the true stress and s_F the plastic strain at fracture.

For n the most likely hypothesis will be to take $n = \frac{1}{2}$, as for Griffith cracks, but obviously other values are possible. The value of N, also, can only be guessed; it may be bound up with the dislocations in cold-worked metals, which are of the order 10^{-5} cm. apart. Fortunately, since N occurs in a logarithm only, the exact value is not important. If $N = 10^{15}$, then, if V does not differ by more than a few powers of 10 from 1 c.c.,

$$\gamma \sim 160 P_2/P_F(1+s_F).$$

Some values of P_2, P_F, s_F have been deduced below from curves given by Körber & Rohland (1924):

	reduction in area	true U.T.S. P_F (kg./mm.^{2})	P_2	γ
iron	0·83	54	34	20
steel 0·1C	0·70	70	42	42
0·25C	0·63	80	45	53
0·45C	0·57	82	38	67

For mild steel, then, according to the formulae of § 2, a bomb of the design suggested there would give fragments of average length 0·6 in.

Referring to formula (5), and assuming that $\log N$ is about the same for all materials, it is seen that the average fragment length is proportional to the following term which depends on the properties of the material:

$$P_F \sqrt{\left(\frac{1+s_F}{\rho P_2}\right)}.$$

Thus a material with a high-stress P_F at fracture gives large fragments, and a high ductility (s_F is the strain at fracture) also gives large fragments. A rapid rate of hardening near the fracture point (i.e. a large value of P_2) will lead to small fragments. Since most metals harden much more rapidly in the initial stages of cold work, low ductility will for this reason also lead to small fragments.

This work was carried out while the author was working in the Ministry of Supply and he wishes to thank the Director-General for Scientific Research and Development for permission to publish it.

REFERENCE

Körber, F. & Rohland, W. 1924 *Mitt. K.-Wilh.-Inst. Eisenforsch.* 5, 55. (See also Körber, F. & Krisch, A. 1939 *Handbuch der Werkstoffprüfung*, 2, 44.)

On a method of increasing contrast in microscopy

By Sir Thomas Merton, Treas.R.S.

(*Received* 29 *August* 1946.—*Read* 14 *November* 1946)

[Plate 7]

When an object is viewed under the microscope, its visibility depends on its opacity, its colour and the extent to which it differs in its refractive index from the surrounding medium. There are a number of methods by which the contrast of a transparent object, whose refractive index differs slightly from the surrounding medium, can be enhanced. These include dark-ground illumination and, in the special case of birefringent materials, the use of polarized light. The most recent and successful addition to these methods is the phase-contrast technique which owes its inception to Zernike (1934) and has since been the subject of investigation by a number of other workers.

It has occurred to me that a powerful method of increasing contrast in the case of transparent substances, in particular certain specimens of biological interest, would be provided by placing the object under investigation between two half-silvered or half-metallized surfaces which would thus constitute an interferometer. Our knowledge of the interference phenomena which occur under such conditions has been greatly enlarged in a recent series of papers by Tolansky (1943 *a*, *b*, 1944 *a*, *b*, 1945, 1946).

For the present purpose it may be noted that a large number of effective reflexions between the plates, giving rise to very narrow fringes, is not essential. The first experiments were carried out by half-silvering two plates of glass (which were not optically flat) so that the amount of light reflected from a surface was about 40 to 50 %. A small amount of pyroxyline was dissolved in amyl acetate and a fine spray was allowed to fall on one of the silvered surfaces by the use of an atomizer. When the amyl acetate had evaporated, the surface of the plate was covered with a large number of thin flat disks of pyroxyline. A drop of water was put on the plate and the second plate was firmly pressed down, the water which was driven out being removed with a piece of filter paper. The plates were put on the stage of a microscope and examined with a low-power objective, the source of light being a mercury arc lamp. The two yellow rays were absorbed by a thick sheet of didymium glass, the visible rays being in effect limited to the green line at $\lambda 5461$ A and the violet line at $\lambda 4359$ A. If the two plates had been optically flat and parallel to one another the field would (in the absence of the pyroxyline disks) have been uniform in colour, but as they were not perfectly flat the field was crossed by a number of broad bands which formed a kind of contour map showing the distance between the plates. With glass of reasonably good quality these bands were for the most part broad in comparison with the objects under investigation. At many parts of the field the colour alternates from green to violet, whilst at some points there was an approximate coincidence between the maxima and somewhat neutral bands with intervening dark spaces were seen. The pyroxyline disks presented a remarkable appearance, being in general conspicuously different in colour or intensity to the particular part of the background against which they were seen. A disk might be seen as a green circle on a violet background or vice versa, and the thickening which occurred at the edges of the disks was manifest in many cases from the appearance of a narrow circle of different intensity or colour. These pyroxyline disks, when examined in the same way on unsilvered glass, were difficult objects to see. At this stage of the investigation the use of silvered surfaces was abandoned. Whilst these half-silvered surfaces are of the highest efficiency from an optical point of view they were so easily scratched or otherwise damaged that it was found preferable to use half-platinized surfaces in which the platinum had been 'burnt' into the glass and which were not impaired by any ordinary usage. The half-platinized surfaces were prepared by a slight modification of the method described by Rheinberg (1920). The plates were flowed over with a solution consisting of chloroplatinic acid 2·1 g., collodion 2 g., ethyl alcohol 77 c.c., acetone 33 c.c., which was compounded by dissolving the chloroplatinic acid in the alcohol and the collodion in the acetone and then mixing the two solutions. When the plate had been allowed to dry in a warm place it was put on a flat plate of stainless steel about ½ in. thick, a foot long and 4 in. in width. One end of the plate was heated by means of a bunsen burner to a temperature estimated at 500 to 600° C and the plate gradually pushed from the cold end to the hot end until the collodion had burned off and the platinum was left as a mirror on the glass. After cooling, the plate was flowed over with a similar solution in which about 0·2 % of bismuth chloride with a few drops of hydrochloric acid was substituted for the chloroplatinic acid. On burning off the collodion a slight bloom was left on the platinum, but this bloom was removed by waving the bunsen flame over the surface until it disappeared. With a little practice the thinnest microscope cover-glasses (about 0·17 mm. in thickness) can be successfully coated without distortion and the surfaces are remarkably durable and permanent.

Certain difficulties arise when objectives of short focus and large N.A. are used. Since the rays pass through the plates in a wide cone, and since the difference of path of the successively reflected beams is $2d\mu \cos \theta$, where d is the distance between the plates, μ the refractive index of the medium and θ the angle which the rays make with the normal, it follows that the colours are diluted to a degree depending on the angle of the cone subtended and the distance between the plates. Moreover, since the interferometer constitutes an angular filter it follows that the objective cannot be uniformly illuminated, with the consequent loss of some resolving power. There are several methods which might be adopted to mitigate this difficulty, and it is proposed to make further experiments in this direction; possibly by the use of an annular stop or a zone plate immediately below the substage condenser.

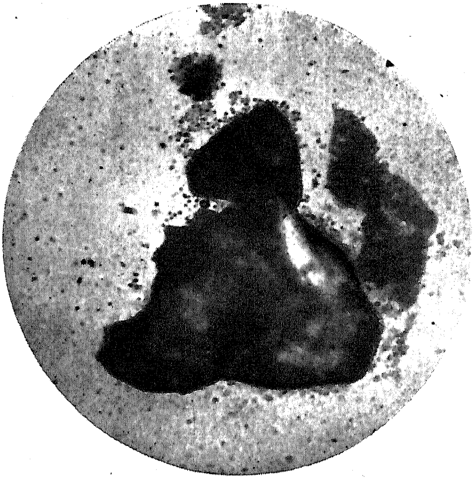


FIGURE 1

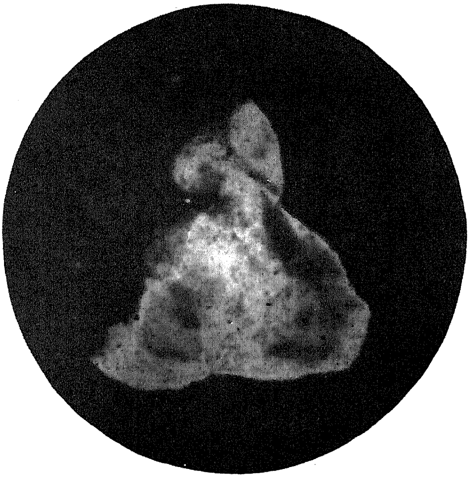


FIGURE 2

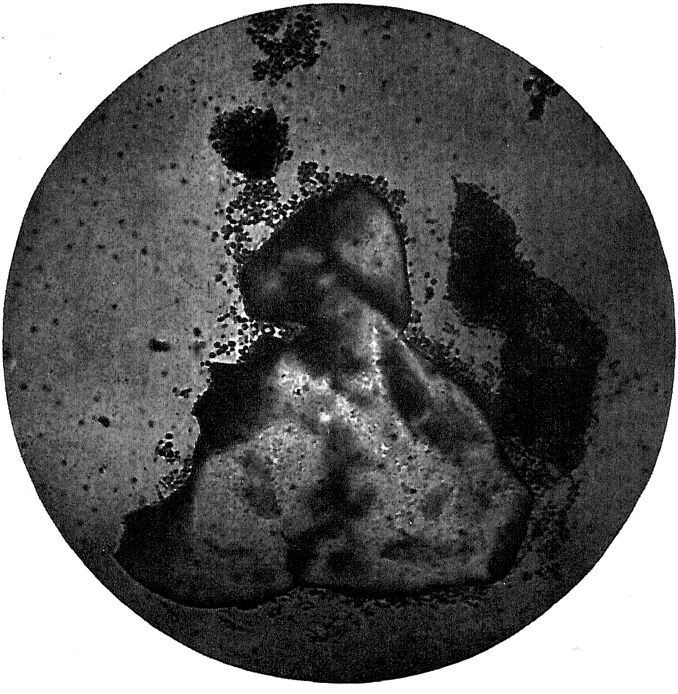


FIGURE 3

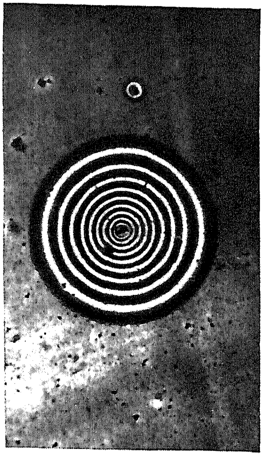


FIGURE 4

I have found, however, that the following rather inelegant and unorthodox but very simple method gives excellent results. A piece of ground glass is placed on the stage of the microscope with the ground surface uppermost, or alternatively the underside of the platinized slide is ground. The image of a mercury lamp after traversing the didymium glass filter and after reflexion from the mirror of the microscope is brought to a focus on the ground-glass surface by means of a lens of $1\frac{5}{8}$ in. focus working at $f/3\cdot1$. This results in the objective being illuminated in a manner determined by the polar curve of the ground glass, the greatest intensity being along the optical axis. This method has been used with objectives down to 4 mm. focus, and highly coloured fringes with good definition have been obtained.

As a test object I have, following Burch & Stock (1942) used wet epithelial cells from the dorsum of the tongue. The cells are easily detached from the surface of the tongue, and a drop of the saliva containing them (with as few bubbles as possible) is laid on the platinized slide. The cover-glass is laid down on the drop and gently squeezed down under a piece of filter paper which removes the liquid which exudes from the sides. Many of these cells show very poor contrast when viewed in the usual manner by transmitted light, and indeed some of them can only be found with difficulty, but when seen between the half-platinized plates they are even more striking than the pyroxyline disks; the structure can be seen in great detail, and variations in thickness are clearly shown as changes in colour or intensity. I have not yet had the opportunity to try these methods with a 2 mm. oil-immersion objective, but there seems to be no reason to suppose that good results will not be obtained.

For practical purposes the method seems to be limited to specimens not exceeding a few microns in thickness. For greater thicknesses the loss either of contrast or of resolving power seems prohibitive. For the highest degree of contrast it would be necessary to illuminate the specimen with parallel or nearly parallel light. This would reduce the resolving power to a degree which would seriously limit the application of the method. The distribution of intensity over the objective resulting from the use of the ground glass provides a useful compromise between loss of contrast and loss of resolving power: but it is essential that the distance between the semi-reflecting surfaces should be small.

There is another method which may be useful occasionally for the study of very small particles in the case in which they can be examined in a dry state.

When a film of cellulose nitrate or cellulose acetate is laid on a plate of glass with a film of water between the cellulosic film and the glass, the water will evaporate through the film, which is left in optical contact with the glass. Owing to the contraction of the film on drying it is necessary to paint over the junction of the film and the bare glass with a solution of gelatine which prevents the film from curling up and detaching itself from the glass on drying.

It is possible to prepare films of silvered or platinized cellulosic material by silvering or platinizing a sheet of glass (in the case of platinizing the treatment with bismuth must be omitted) and by pouring over the surface, for example, a $2\frac{1}{2}$ % solution of cellulose nitrate in amyl acetate. When the solvent has evaporated the film can be detached from the glass and usually carries the silver or platinum with it. In a similar way metallized films of cellulose acetate can be prepared by the use of a 15% solution of cellulose acetate in acetone which is spread over the silvered or platinized glass with a steel straight-edge with copper wires of 30 to 40 s.w.g. as distance pieces along the edges of the glass.

If a piece of half-metallized film prepared in this way is laid on a half-metallized sheet of glass with a film of water between the surfaces and the water is allowed to evaporate, the two metallic surfaces are left in optical contact. If the water between the film and the glass contains any solid particles the cellulose film is raised by these particles on drying, and on examining the plate through a microscope with a mercury arc as a source of illumination, each particle is seen to be surrounded by a concentric series of interference fringes. By counting the number of rings and estimating the fraction of a ring, the thickness of the particle can at once be determined, N rings denoting a thickness $\frac{1}{2}N\lambda$.

In the same way a fragment of a fibre will be flanked by a series of linear fringes which follow the contour of the fibre and which provide a measure of its thickness.

The fringes which are produced in this way differ conspicuously from the Newton fringes which are seen when a half-silvered convex surface is laid on a half-silvered plate. Owing to the manner in which the cellulosic film is distorted by the particle, the fringes are closest together at the centre of the system and get wider and wider apart until zero difference of path is reached at a distance from the centre at which the two metallic films pass into metallic contact.

It will be seen, however, that the exact law governing the separation of the rings is irrelevant in so far as the determination of the thickness of the particle is concerned, since it is only their number and an estimate of a fraction of a ring which is required. If the mercury line $\lambda = 4358$ A is used the first maximum would denote a thickness of 0·00022 mm., whilst a particle which raised the film to a height such that the first minimum fell on the centre of the particle would have a thickness of half this value or 0·00011 mm. The method provides a simple means of measuring very small particles and fibres with a considerable degree of precision.

There is, however, a further consideration which may prove to be of greater importance. A perceptible darkening can be observed at a separation which is much smaller than the separation corresponding to the first minimum, and it seems possible that with a sufficiently powerful source of illumination and films of high reflecting power particles as small as 10^{-5} mm. might be detected.

It is hoped that these methods may find some useful application in the biological field.

Note added December 1946. Since this paper has gone to the printers my attention has been called by Dr C. R. Burch, F.R.S., to the work of A. M. Frederikse (*Acta Brev. Neerl.* **111**, 8/9 (finit 10. x. 1933)) in which the interferometer has been applied to the study of biological specimens. Frederikse's paper covers a part of the work described in this communication and I am glad to be able to call attention to his work.

REFERENCES

Burch, C. R. & Stock, J. P. P. 1942 *J. Sci. Instrum.* **19**, 71.
Rheinberg, J. 1920 *Discussion on making reflecting surfaces.* London: Physical Society.
Tolansky, S. 1943*a* *Nature*, **152**, 722.
Tolansky, S. 1943*b* *Phil. Mag.* **34**, 555.
Tolansky, S. 1944*a* *Nature*, **153**, 195.
Tolansky, S. 1944*b* *Phil. Mag.* **35**, 120.
Tolansky, S. 1945 *Proc. Roy. Soc.* A, **184**, 41.
Tolansky, S. 1946 *Proc. Roy. Soc.* A, **186**, 261.
Zernike, F. 1934 *Mon. Not. R. Astr. Soc.* **94**, 377.

DESCRIPTION OF PLATE 7

FIGURE 1. Epithelial cells from the tongue with groups of bacteria between half-platinized surfaces photographed in the light of the two mercury yellow lines. These cells are colourless and transparent (imperfect focusing).

FIGURE 2. The same cells photographed in mercury green light (bacteria nearly invisible).

FIGURE 3. The same cells photographed in mercury violet light (× 450).

FIGURE 4. Two particles of rouge, showing ten and one complete rings respectively, between a platinized pellicle of cellulose acetate and a platinized plate.

The spectra of flames containing oxides of sulphur

BY A. G. GAYDON, D.SC., *Chemical Engineering Department, Imperial College,*
AND G. WHITTINGHAM, PH.D., *British Coal Utilization Research Association*

(*Communicated by Sir Alfred Egerton, F.R.S.—Received* 30 *May* 1946)

[Plates 8 and 9]

The spectra and characteristics of hydrocarbon and related flames containing small amounts of SO_2 and SO_3 have been studied; observations have been made with a view to elucidating the mechanism of carbon formation in flames.

Band systems due to S_2, SO, CS and SH occur; this is the first record of the SH band in emission. The mechanism of the formation of these radicals is discussed. The outer cones of flames containing SO_2 show strong ultra-violet continuous emission, and this is provisionally attributed to direct association between SO_2 and atomic oxygen to give SO_3. Failure to obtain luminous reaction between SO_2 and atomic oxygen in a discharge tube at room temperature indicates that this reaction requires an activation energy. The formation of SO_3 cannot be used as a quantitative test for atomic oxygen in flames.

SO_3 increases carbon deposition in hydrocarbon flames and it is considered that it can induce chain reactions leading to the polymerization and decomposition of hydrocarbons. These chain reactions are maintained by free radicals which may be produced either by direct reaction of sulphur trioxide with hydrocarbons or indirectly through the formation of peroxides. SO_2 and H_2S, which reduce carbon deposition, either inhibit the chain processes or remove carbon as soon as it is formed.

INTRODUCTION

The formation of sulphur oxides during the combustion of fuels containing sulphur compounds is a subject of technical importance, and the oxidation of sulphur dioxide in diffusion flames of CO-air, H_2-air, and CH_4-air has been studied by Dooley & Whittingham (1946) as part of an investigation on the role of sulphur compounds in combustion sponsored by the Boiler Availability Committee. For a given concentration of SO_2 the highest conversion to SO_3 was in the CO-air flame, and the luminous, smoky flame of CH_4-air was the least effective in this respect. It was considered that the association reaction

$$SO_2 + O = SO_3 + 81\,\text{kcal.}$$

was responsible for the oxidation.

The role of atomic oxygen in combustion reactions has been discussed by Gaydon (1944), and a sensitive technique was described which appears to give direct evidence of the existence of O atoms in various flames. It was shown that the yellowish-green continuum emitted by flames to which small amounts of nitric oxide had been added was identical with the spectrum of the air afterglow and could be attributed to the reaction of NO with atomic oxygen. Application of this technique showed that there was a relatively high concentration of free oxygen atoms in CO flames, but during the initial stages of CH_4 combustion the concentration was low. These results were supported by the extent to which SO_2 was oxidized in CO-air and CH_4-air flames, and in the discussion on the paper by Dooley & Whittingham one of us (A. G. G.) suggested that a continuous emission might be detected in flames containing oxides of sulphur due to the reaction

$$SO_2 + O = SO_3 + 81\,\text{kcal.},$$

the emission lying at wave-lengths rather longer than 3500 A.

The luminosity of coal-gas flames has been found to be markedly influenced by the presence of added sulphur oxides. It was shown (Whittingham 1945) that in the presence of about 4 % SO_2 the luminosity of an under-aerated coal-gas flame was greatly reduced, and the resultant flame was strongly reminiscent of burning sulphur. On the other hand, the addition of about 0·1 % by volume of SO_3 rendered a Bunsen flame luminous. Although a brief spectroscopic examination of the flames was carried out then, a more complete investigation seemed desirable, as it was thought that the results might have a bearing on the mechanism of carbon deposition from flames, a process about which comparatively little is known.

There appears to be little published information on the spectra of flames containing oxides of sulphur, although the flame spectra of other sulphur compounds have received some attention. Fowler & Vaidya (1931) studied the flame spectra of sulphur, hydrogen sulphide and carbon disulphide, and the principal features of these were a strong band system due to sulphur, S_2, and a weaker band system of sulphur monoxide, SO. The cool-flame spectrum of CS_2 was investigated by Emeléus

(1926), and, in addition to the above band systems, weak bands due to the CS radical were found to be present.

In the present investigation a study has been made of the influence of added sulphur oxides on the spectra of CO, H_2, CH_4, C_2H_4, CH_3OH and coal-gas flames, and the results, taken in conjunction with those obtained from the chemical analysis of the products of combustion of flames containing added SO_2, afford useful information on the role of sulphur compounds in combustion reactions. Some preliminary observations of the spectra of coal-gas flames containing H_2S have also been made for comparison.

EXPERIMENTAL

Sulphur dioxide was obtained from a syphon, and hydrogen, carbon monoxide, methane and ethylene were obtained from cylinders and were used without further purification or drying. Coal gas was burnt in a Bunsen burner and the other gases burnt at a quartz jet. Small amounts of SO_3 were introduced into the flames by passing a slow stream of air through oleum and allowing this to mix with the main gas stream.

The flame spectra were photographed on a Hilger small quartz instrument, this giving adequate dispersion for the purposes of this work, and a fairly wide slit was usually found adequate.

SPECTRA AND OTHER CHARACTERISTICS OF FLAMES CONTAINING SO_2

Coal gas. An ordinary fully aerated Bunsen flame shows the usual C_2, CH and OH bands in the blue-green inner cone, and the OH band and weak CO flame spectrum in the pale blue-violet outer cone. When sulphur dioxide up to about 4 % by volume is introduced into a flame of this type the inner cone elongates slightly, due to a reduction in the flame speed, and becomes more blue-violet in colour, while the outer cone becomes much brighter. The spectrum of the inner cone now shows fairly strong SO bands ($\lambda\lambda$ 2877, 2968, 3164, 3271, 3383, 3502, etc.) and the CS bands around 2575 A, while the C_2 and CH bands are weakened; the C_2 are probably weakened to a greater extent than the CH (plate 8*a* and *b*). The spectrum of the outer cone is continuous apart from the OH bands, the continuum extending from the blue to well down in the ultra-violet.

Analysis of the interconal gases in this type of flame has shown that about 25 % of the original SO_2 is reduced to H_2S and another 10 % to free sulphur. Small amounts of CS_2 and COS are also produced, but the amount of SO_3 formed in the inner cone is negligible.

An under-aerated Bunsen flame is, of course, luminous, due to deposition of solid carbon, and the inner and outer cones are not clearly differentiated. If a Bunsen is adjusted so that the top of the flame is just luminous and then a little SO_2 is added to the gas supply the luminosity, due to solid carbon, disappears completely. The spectrum of a flame of this type containing SO_2 then shows fairly strong CS bands and some S_2 bands in the violet and near ultra-violet and perhaps a trace of the SO bands. There is also a weak double-headed band, very similar in general appearance to the OH band, with heads at about 3235 and 3278 A; this is presumably to be identified with the SH band which has been reported in absorption by Lewis & White (1939) and has heads at 3236·6 and 3279·1 A. This band has never been recorded in emission before, despite many attempts to observe it. It does not occur in the fully aerated Bunsen flame, in the flame of hydrogen burning in air when SO_2 is present, nor in a flame of burning H_2S. Previous failures to obtain this band in emission have been attributed to predissociation. These weak bands in the under-aerated flame are superposed on a relatively strong continuous emission extending from the blue to around 2500 A, but the continuum in the visible due to incandescent carbon particles is suppressed by the addition of SO_2 (see plate 8*c* and *d*).

Hydrogen. A simple flame of hydrogen burning in air at a quartz jet is practically non-luminous, the spectrum showing only the strong ultra-violet OH bands. On addition of a few per cent of SO_2 the flame becomes blue-violet in colour and a weak blue inner cone is faintly visible. The spectrum of this faint inner cone shows fairly strong S_2 bands (see plate 8*e*), but the SO bands have not been observed in this source. The outer cone of this flame (plate 8*f*) shows continuous emission and the OH bands; the maximum of the continuum is around 3400 A, and it extends from the blue to around 2800 A (the limits are not, of course, sharp, and can be considerably extended by increasing the exposure time). An aerated hydrogen flame shows a similar spectrum and again there is no sign of SO band structure.

Carbon monoxide. The normal blue flame of carbon monoxide burning in air shows a continuous spectrum with maximum around 3700 A; the continuum extends from the green to about 2800 A, or further with long exposures. A weak band structure is superposed on this, and with undried gases the OH band at 3064 A is always strong.

On addition of SO_2 the flame becomes rather more violet in colour and there is some alteration in the distribution of the continuous spectrum. The maximum appears to shift slightly towards shorter wave-lengths and the ultra-violet end is definitely extended. The OH band is practically eliminated by introduction of SO_2. Plate 8, *g* and *h*, shows the comparison between the normal CO flame and that with SO_2 added. With SO_2 present the spectrum appears to be a summation of the continuum due to the CO flame and that due to SO_2 as shown in the H_2+SO_2 flame.

Methane, CH_4. The effect of SO_2 on methane is very similar to that on coal gas. In the inner cone of the fully aerated flame, bands of SO and CS appear but not S_2, and the C_2 and CH bands are weakened (see plate 9*b*). With the under-aerated flame the SO_2 has again a marked effect in reducing carbon formation.

Ethylene, C_2H_4. The observations were again similar to those with coal gas. Introduction of SO_2 into the aerated flame lengthened the inner cone slightly, weakened the C_2 and CH bands and produced CS and SO bands and weak SH. In this case the S_2 bands are fairly well developed in the outer cone. The hydrocarbon flame bands (HCO ?) which are relatively strong in this flame are probably suppressed by the addition of SO_2, but it is rather difficult to be certain of this as the strong

SO bands fall in the same region and mask the HCO bands when SO_2 is introduced (see plate 9*c* and *d*).

Methyl alcohol, CH_3OH. A methyl alcohol-air flame was maintained by bubbling air through a gas washing bottle fitted with a sintered disk and containing methyl alcohol warmed to about 35° C, and burning the saturated vapour at a Pyrex jet. The normal flame showed chiefly OH and CO flame spectrum with some CH and weak C_2 bands. The addition of about 4 % SO_2 considerably strengthened the continuous emission and shifted it towards shorter wave-lengths and produced weak SO bands. The OH, CH and C_2 bands were weakened.

THE EFFECT OF H_2S ON COAL-GAS FLAMES

When hydrogen sulphide is added to coal gas burning in a Bunsen burner the effects are generally very similar to those observed when SO_2 is added. With a luminous under-aerated flame the amount of carbon formed is reduced by the addition of H_2S. The spectrum of an under-aerated coal-gas flame, the luminosity of which has been removed by means of H_2S, is generally similar to that of the corresponding flame with SO_2. There is considerable continuous emission, upon which faint bands of S_2, SH and CS are superposed. The S_2 bands are probably stronger than in the corresponding flame containing SO_2 while the CS and SH bands are less strong, the latter being barely discernible.

With a fully aerated, i.e. Bunsen, flame the addition of H_2S lengthens the inner cone, due to reduction in the flame speed, and the colour changes from blue-green to blue-violet. CS, SO and SH bands appear and C_2 and CH are reduced in intensity. By comparison with the fully aerated flame containing SO_2, the CS band produced by H_2S is relatively strong. The SH is largely masked by the SO bands but the first head at 3236 A is clearly present; comparison of plates 8*b* and 9*e* shows the presence of the head when H_2S is added but not with SO_2.

SPECTRA AND OTHER CHARACTERISTICS OF FLAMES CONTAINING SO_3

Coal gas. The most striking effect of the introduction of small amounts (of the order 0·1 % by volume) of sulphur trioxide into a Bunsen flame is the change from a clear to a luminous flame. This is most marked when the flame is initially adjusted to be just non-luminous, but formation of incandescent carbon particles occurs even with a fairly highly aerated flame. Separation of the two cones of the flame in a Smithell's separator shows that the carbon formation occurs in the inner cone, and that the inner cone is lengthened slightly when SO_3 is added. The spectrum of this inner cone shows, in addition to the continuum in the visible due to the hot particles, some SO bands and a trace of CS. The C_2 and CH bands are somewhat weakened by the introduction of SO_3, but the effect is less marked than with SO_2; the OH bands are not appreciably reduced in intensity (see plate 9*f* and *g*). There is a slight increase in the intensity of the ultra-violet continuous emission from the outer cone, but this is very much less striking than with SO_2.

Hydrogen. The addition of SO_3 to a hydrogen flame does not produce any marked change in the appearance of the flame. It becomes slightly more violet in colour, and the spectrum shows weak S_2 bands.

Carbon monoxide. The addition of SO_3 produces little effect on the carbon monoxide flame. There may be some strengthening of the short-wave end of the continuous emission spectrum, but the effect is less marked than when SO_2 is added. There is no very obvious reduction in the strength of the OH band, as with SO_2, and no new bands appear in the spectrum; the concentration of SO_3 was, however, much less than that of SO_2.

THE REACTION BETWEEN SULPHUR DIOXIDE AND ATOMIC OXYGEN

It has been shown by Dooley & Whittingham (1946) that the oxidation of SO_2 to SO_3 in flames is probably due to reaction between atomic oxygen and the SO_2. The reaction seems to bear some resemblance to that between nitric oxide and atomic oxygen to form nitrogen peroxide. This latter reaction appears to occur as a direct association at room temperature and is accompanied by the emission of a continuous spectrum with maximum in the yellow-green. It therefore appeared of interest to see whether there was any reaction between atomic oxygen and sulphur dioxide at room temperature.

A simple discharge tube of the same general type as that used by Gaydon (1944) for studying the NO-O reaction was used. Commercial oxygen passed through the discharge at a pressure of a few mm. of mercury and then mixed with a stream of SO_2 issuing from a narrow jet. A liquid-air trap served to protect the pump, and the tube was observed in a completely darkened room. The discharge through the oxygen is not very luminous, but quite a well-developed yellow-green air afterglow (from traces of nitrogen in the oxygen) was observed. The addition of a little SO_2 did not produce any further glow or have any marked effect on the air afterglow. A fast stream of SO_2 did appear partially to suppress the air afterglow, although this effect may have been due, at least in part, to the raising of the gas pressure in the discharge tube by the addition of SO_2, thus rendering the conditions less favourable for the production of an afterglow of long duration. Slight heating of the tube at the point of entry of the SO_2 failed to produce any new glow, but did slightly suppress the air afterglow.

Thus it is clear that there is no evidence for a luminous reaction between SO_2 and atomic oxygen at room temperature. It was, however, noticed that droplets were formed on the wall of the tube and tests showed these to consist of sulphuric acid. This presumably indicated formation of SO_3 either by three-body collisions or by direct association without emission of light; the former seems more probable, but the SO_3 molecule has six vibrational degrees of freedom and the possibility of the excess energy being stored in these, thus permitting direct two-body association without emission of light, cannot be entirely excluded. The formation of sulphuric acid at room temperature agrees with Geib & Harteck's (1934) observation that in the presence of atomic oxygen about 8 % of sulphur dioxide was oxidized to the trioxide at liquid-air temperatures.

However, although there is no luminous reaction between SO_2 and atomic oxygen in the cold, it is possible that such a reaction may take place at high temperatures. It seems that in a collision between NO and O the particles may run together on a stable potential surface (i.e. one with a minimum), and that before the particles swing apart there is a small chance that a transition may occur to the ground state of NO_2 with the emission of radiation and stabilization to NO_2. In the case of SO_2 and O the collision may occur on a repulsive potential surface and the particle will then approach sufficiently closely to give an appreciable change of a transition to the ground state of SO_3 with emission of radiation only if they collide with fairly high kinetic energy.

The outer cones of flames containing SO_2 emit strong continuous radiation, the maximum intensity being around 3400 A and the continuum extending well towards shorter wave-lengths. The reaction $SO_2 + O = SO_3$ yields about 81 kcal., which corresponds to a wave-length of around 3500 A. Clearly radiation shorter than this could be emitted only if the colliding particles possessed kinetic energy. It is noteworthy that the flame in which the continuous emission is strongest, the CO flame, is that in which most oxidation of SO_2 occurs. This flame is also slightly hotter than the hydrogen flame, and the continuum definitely extends further to short wave-lengths ($\sim$2500 A) in the CO flame than in the hydrogen flame ($\sim$2800 A).

Thus it is possible to make out quite a good case for the process responsible for the emission of the continuum from the outer cones of flames being identified with $SO_2 + O = SO_3 + h\nu$ reaction. Against this it should be pointed out that discharge-tube sources containing oxides of sulphur also emit strong continuous radiation, and that the afterglow of SO_2 (Gaydon 1934) also emits a strong continuum which was believed to be due to the association process $SO + O = SO_2$. On the whole it seems likely that the continuum is due to reaction between SO_2 and O to form SO_3, but that the formation of SO_3 in flames is not necessarily dependent on this reaction with emission of radiation, as there is evidence that combination can also occur in the cold, either as the result of three-body collisions or as direct association without radiation. Assuming that the continuum is emitted by this reaction, its strength would be expected to increase with rising temperature, so that, even apart from the possibility of three-body collisions forming SO_3, one could hardly expect the rate of formation of SO_3 in flames to lead to a strictly quantitative measure of the concentration of atomic oxygen.

DISCUSSION OF THE BAND SPECTRA

In discussing the spectroscopic observations it must first be remembered that the observations are of emission from electronically excited molecules or radicals. This emission may, in part, be thermal in origin, but there is good reason to believe that chemiluminescence is also important in determining the emission from flames. Thus suppression of a band system, as, for example, the OH band when SO_2 is introduced into a flame of moist CO, might mean removal of the radical by some chemical reaction, e.g.

$$SO_2 + OH = SO_3 + H - 20\,\text{kcal.}, \tag{1}$$

but it might equally well be explained as deactivation of the electronically excited radical without chemical action. Production of the band spectrum of a molecule or radical may, similarly, be due to its formation in an excited state (chemiluminescence) or to its formation in the normal state followed by thermal excitation.

In order to explain the various conceivable reactions which might account for the formation or removal of molecular species or radicals some knowledge of the thermochemistry of the process is desirable. Unfortunately, the thermochemistry of many of the less stable compounds of sulphur is still in doubt, e.g. the dissociation energies of S_2, and CS. In the following discussion the values recommended by Gaydon (1947) for the heats of formation (from free atoms) are used. These are (expressed in kcal.): C_2 83?, CH 80, CO 256, H_2 103, O_2 117, OH 101, S_2 101, SO 119, CS 166, H_2O 219, H_2O_2 253, CO_2 381, SO_2 252, SO_3 333, SH 70?. The energies of excitation used are: CH* 67, OH* 93, C_2* 55, CS* 104, SO* 103, SH* 88 (throughout the discussion an asterisk denotes an electronically excited molecule).

One of the most striking experimental observations is the effect of SO_2 on hydrocarbon flames. Carbon deposition in an under-aerated flame is suppressed and CS bands are emitted with a fairly high intensity. In a fully aerated flame the intensities of the C_2 and CH bands in the inner cone are reduced and bands due to SO and CS appear, the latter being weaker than in the under-aerated flame.

Among the possible reactions leading to the formation of CS and SO are:

$$C + SO_2 = CO + SO^* + 20\,\text{kcal.}, \tag{2}$$

$$C + SO_2 = CS + O_2 + 31\,\text{kcal.}, \tag{3}$$

$$C_2 + SO_2 = CS^* + CO_2 + 108\,\text{kcal.}, \tag{4}$$

$$C_2^* + SO_2 = CS^* + CO_2 + 163\,\text{kcal.}, \tag{5}$$

$$C_2 + SO_2 = C + CO + SO + 40\,\text{kcal.}, \tag{6}$$

$$C_2^* + SO_2 = C + CO + SO^* - 8\,\text{kcal.}, \tag{7}$$

$$CH^* + SO_2 = CS + H + O_2 + 18\,\text{kcal.}, \tag{8}$$

$$CH^* + SO_2 = CS + HO_2 \pm ?, \tag{9}$$

$$CH^* + SO_2 = CO + H + SO + 110\,\text{kcal.}, \tag{10}$$

$$CH^* + SO_2 = CO + H + SO^* + 7\,\text{kcal.}, \tag{11}$$

$$CH + SO_2 = CO + H + SO + 43\,\text{kcal.} \tag{12}$$

Thus the emission of the CS band from hydrocarbon flames is more readily accounted for in terms of reactions (4) and (5), as some degree of thermal excitation would be required in (3). The only reactions producing SO* either require free carbon atoms or else are slightly endothermic. The fact that the SO bands are strong in the hot inner cones of fully aerated flames where the C_2 and CH bands are intense, suggests that the SO bands are due to reactions (7) and (11) rather than (2).

Reactions such as (6) and (12), which form normal SO, followed by thermal excitation, seem less likely as the SO bands are largely restricted to the reaction zone of the inner cone.

The reduction in the strength of the OH bands when SO_2 is present may be explained by reaction (1). It is of interest to note here that all the bimolecular reactions which lead to the formation of SO or SO* involve the formation of CO as well. With the high value for the dissociation energy of CO used here (256 kcal.) this presents no difficulty, but if a lower value for the dissociation energy were used, all these reactions would liberate much less energy, becoming mostly very endothermic. This constitutes an indirect argument in favour of the high value for the dissociation energy of CO.

The occurrence of the weak SH bands in the under-aerated flames of hydrocarbons is interesting, especially because the band is not known to occur in the flame of H_2+SO_2 or in the simple flame of burning H_2S. It is thus improbable that it is due to reaction between H, H_2, or OH, and SO_2. It may be formed by some subsequent reaction of either CS or SO formed during the combustion of hydrocarbons in the presence of SO_2. Possible reactions are:

$$SO^*+H_2 = OH+SH^*-\sim 36\,\text{kcal.}, \quad (13)$$

$$SO+CH = CO+SH^*+\sim 39\,\text{kcal.}, \quad (14)$$

$$CS+OH = CO+SH^*-\sim 29\,\text{kcal.}, \quad (15)$$

$$CS^*+OH = CO+SH^*+\sim 75\,\text{kcal.}, \quad (16)$$

$$CS^*+CH^* = C_2+SH^*-\sim 10\,\text{kcal.} \quad (17)$$

Since the SH band occurs under the conditions in which the CS bands are strongest, i.e. an under-aerated coal-gas flame containing SO_2 and an aerated flame containing H_2S, and the SO bands are not present with appreciable strength, reactions (15) or (16) seem the most likely.

The origin of the S_2 bands in the hydrogen flame is not entirely clear. They are known to be a persistent impurity in certain flames containing hydrogen and appear especially if oxygen is burnt in an atmosphere of hydrogen containing a trace of sulphur, or if an impure coal-gas flame is brought into contact with a cool surface. Brinsley & Stephens (1946) have reported that as little as one part by weight of sulphur in 4×10^6 parts of hydrogen is sufficient to produce the S_2 bands. Most of the bimolecular reactions which could form S_2 in the H_2 flame seem too endothermic to appear likely. Brinsley & Stephens suggest the equilibrium

$$2SO \rightleftharpoons SO_2+\tfrac{1}{2}S_2$$

as responsible for the formation of sulphur.

A suggestion made to us by Mr H. G. Crone is that any sulphur present might be strongly excited by the reaction

$$H+H+S_2 = S_2^*+H_2.$$

There would be sufficient energy from the combination of H atoms to excite S_2, but the margin is not very great, and this may account for the fact that only transitions from a few of the lower vibrational levels of the excited state are observed in this flame. A similar reaction with SO as the third body would not lead to formation of SO^* because the energy is insufficient. SO may be present in H_2+SO_2 flames but in an unexcited state, for example, through

$$H+SO_2 = SO+OH,$$

a reaction which is slightly endothermic.

The introduction of small amounts of SO_3 into hydrocarbon flames, in addition to producing solid carbon, causes slight weakening of the C_2 and CH bands and the appearance of some SO bands. Deactivation of excited molecules and radicals may play some part in reducing the C_2 and CH bands, as SO_3 has a large 'collision diameter', but the SO emission and also this reduction in the strength of the C_2 and CH bands may be attributed to the action of SO_2 arising out of the thermal decomposition or reduction of SO_3.

THE MECHANISM OF CARBON FORMATION IN HYDROGEN FLAMES

Before discussing the role of sulphur oxides in relation to the problem of carbon formation in flames, a subject which has a direct bearing on 'coking' in engines, it is relevant to review previous work on the conditions under which C_2 and CH bands occur in combustion processes.

Not only are C_2 and CH bands given by the flames of hydrocarbons burning at atmospheric pressure, but they also appear when hydrocarbons (other than methane) are subjected to the action of atomic hydrogen (Bonhoeffer & Harteck 1928) and atomic oxygen (Geib & Vaidya 1941) at low pressure and room temperature. It is considered that successive stripping of hydrogen atoms from the hydrocarbon molecule will lead to the formation first of free radicals and finally of CH and C_2. As an alternative possibility the CH may be formed from C_2 by reaction with either OH or H atoms. We have also observed that when chlorine is added to a Bunsen flame there is a very marked strengthening of the C_2 emission, accompanied by an increase in the flame speed, but the effect on the amount of carbon formation is not great; this strengthening of C_2 could again be explained as resulting from stripping of hydrogen atoms from the hydrocarbon by chlorine.

Hsieh & Townend (1939) found that mixtures of hydrocarbons and air burning under reduced pressure gave intense green flames, the colour being due to emission from C_2, and the excitation of CH and C_2 was greatly dependent on the pressure and composition of the mixture. The emission of C_2 was related to the initial stages of hydrocarbon combustion and also depended on the extent to which oxygen was deficient. Smith (1940) followed this work up by observing that there was an abrupt change from this green-flame condition to that of carbon formation in which the C_2 band emission disappeared.

It seems highly probable that the polymerization of C_2 and of hydrocarbons has some bearing on the problem of solid carbon deposition. The high-temperature coefficient of the polymerization of C_2 to solid carbon was remarked on by Klemenc, Wechsberg & Wagner (1934) in their study of the decomposition of carbon suboxide

$$C_3O_2 = C_2 + CO_2.$$

Some observations on the pyrolysis and polymerization of hydrocarbons have been made which seem relevant to this discussion. In the pyrolysis of methane the influence of free radicals in inducing this process has been demonstrated by Hessels, van Krevelen & Waterman (1939). One important fact arising from their work was that the ratio of inductor to methane was very critical in determining the course of the reaction. The polymerization of ethylene is very sensitive to the presence of oxygen (Lenher 1931, 1932), and Storch (1934) has shown that one oxygen molecule leads to the polymerization of 85 molecules of ethylene, thus indicating some chain process in the polymerization. It was also found by Taylor & Jones (1930) that alkyl radicals, formed by decomposition of metal organic compounds, induce polymerization in ethylene.

In Storch's later work (1935) it was found that the polymerization of ethylene was inhibited by small traces of sulphur compounds, particularly H_2S, this effect being attributed to the anti-oxidant effect of the sulphur compounds. It has been suggested that the inducing effect of oxygen might be due to the formation of peroxidic bodies or complexes which subsequently decompose to free radicals which maintain the polymerization chains. With methane there is some evidence, from the work of Waterman & Van Vlodrop (1939) on the influence of H_2S on the pyrolysis, that at temperatures above 1000° C carbon deposition is very small and H_2S is converted to CS_2.

The suppression of carbon formation in luminous flames by the addition of SO_2 or H_2S can be attributed to either interference with the sequence of reactions leading to the formation of carbon aggregates, or reaction with the carbon aggregates once formed, or both. The reaction of solid carbon with SO_2 has been studied by Lepsoe (1940), and is extremely rapid above 900° C; this, and the observation that the CS band is produced strongly when carbon deposition is being suppressed, suggests that removal of solid carbon is at least contributory to the reduction of luminosity when SO_2 or H_2S is added to a flame. On the other hand, the reduction in flame speed, noticeable by the lengthening of the inner cone, indicates that SO_2 and H_2S do also interfere with the combustion reactions. Consequently, removal of radicals by SO_2 or H_2S is probable, and this may reduce polymerization and carbon formation.

The amount of SO_3 required to produce a marked increase in carbon deposition is very small, about 0·1 % by volume. This clearly could not produce a significant change in flame temperature and so could not have a purely thermal effect on polymerization of C_2 or hydrocarbons. Furthermore, since SO_3 is likely to be rapidly destroyed in the inner cone of a Bunsen flame its effect must be confined to the early stages of combustion. This appears to indicate that SO_3 must in some way initiate some chain process the end-product of which is solid carbon. The effect of SO_3 on the C_2 bands is not very marked, and it has already been seen that polymerization of hydrocarbons such as methane and ethylene tends to proceed by a chain mechanism in which free radicals are probably important. Hence it seems most probable that the action of SO_3 is to initiate polymerization of hydrocarbons, the higher polymers of which subsequently decompose thermally to solid carbon. It is uncertain whether the SO_3 first reacts with the hydrocarbon to form a peroxide, which subsequently decomposes to a radical which initiates the polymerization, or whether the SO_3 attacks the hydrocarbon with formation of OH or H_2O, SO_2 and a free radical. It may be noted in passing that the effect of SO_3 in inducing carbon formation is so much greater than that of SO_2 in suppressing it, that the second-order effect of the SO_2 formed by reaction of SO_3 will be practically negligible.

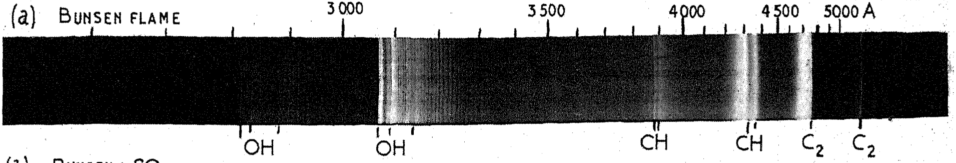
Thus one may conclude that present observations of the influence of SO_2, H_2S and SO_3 on carbon formation in flames of burning hydrocarbons are best explained by assuming that the carbon formation in flames depends on chain reactions, maintained by free radicals, which lead to polymerization of the hydrocarbons, followed by their thermal decomposition. SO_2 and H_2S reduce carbon deposition partly by removing the solid carbon already formed and partly by interfering with the radicals (or the peroxides from which they are formed) which maintain the polymerization chains. The strong effect of very small amounts of SO_3 in inducing carbon deposition may be ascribed to its action in initiating the polymerization of the hydrocarbons either by direct formation of free radicals from the hydrocarbon or by indirect formation of these radicals via a peroxide.

One of us (A. G. G.) is grateful to the Royal Society for the award of a Warren Research Fellowship, and our thanks are due to Sir Alfred Egerton, Dr D. T. A. Townend and Dr D. H. Bangham for their interest in and helpful criticism of the work.

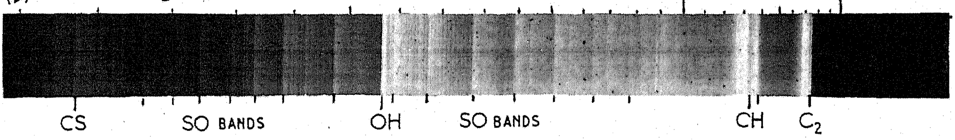
REFERENCES

Bonhoeffer, K. F. & Harteck, P. 1928 *Z. Phys. Chem.* **139**, 64.
Brinsley, F. & Stephens, S. 1946 *Nature*, **157**, 622.
Dooley, A. & Whittingham, G. 1946 *Trans. Faraday Soc.* **42**, 354.
Emeléus, H. J. 1926 *J. Chem. Soc.* p. 2948.
Fowler, A. & Vaidya, W. M. 1931 *Proc. Roy. Soc.* A, **132**, 310.
Gaydon, A. G. 1934 *Proc. Roy. Soc.* A, **146**, 910.
Gaydon, A. G. 1944 *Proc. Roy. Soc.* A, **183**, 111.
Gaydon, A. G. 1947 *Dissociation energies*. London: Chapman and Hall.
Geib, K. H. & Harteck, P. 1934 *Trans. Faraday Soc.* **30**, 134.
Geib, K. H. & Vaidya, W. M. 1941 *Proc. Roy. Soc.* A, **178**, 351.
Hessels, W. J., van Krevelen, D. W. & Waterman, H. I. 1939 *J. Soc. Chem. Ind., Lond.*, **58**, 323.
Hsieh, M. S. & Townend, D. T. A. 1939 *J. Chem. Soc.* p. 332.
Klemenc, A., Wechsberg, R. & Wagner, G. 1934 *Z. Elektrochem.* **40**, 488.

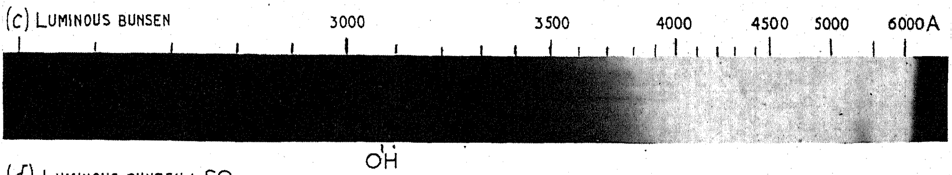
(*a*) BUNSEN FLAME



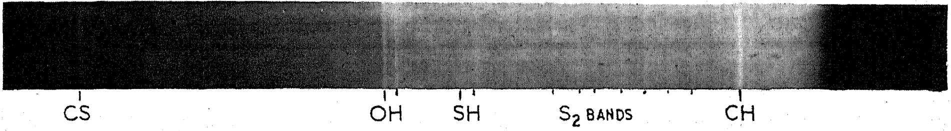
(*b*) BUNSEN + SO_2



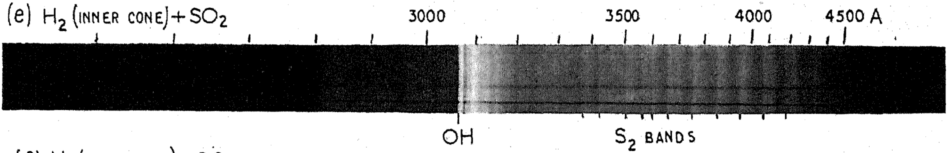
(*c*) LUMINOUS BUNSEN



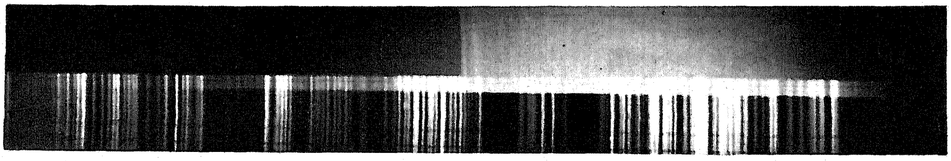
(*d*) LUMINOUS BUNSEN + SO_2



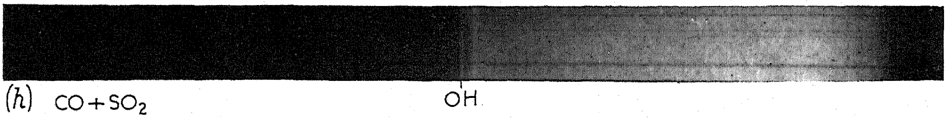
(*e*) H_2 (INNER CONE) + SO_2



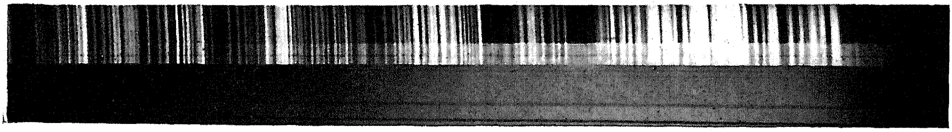
(*f*) H_2 (OUTER CONE) + SO_2

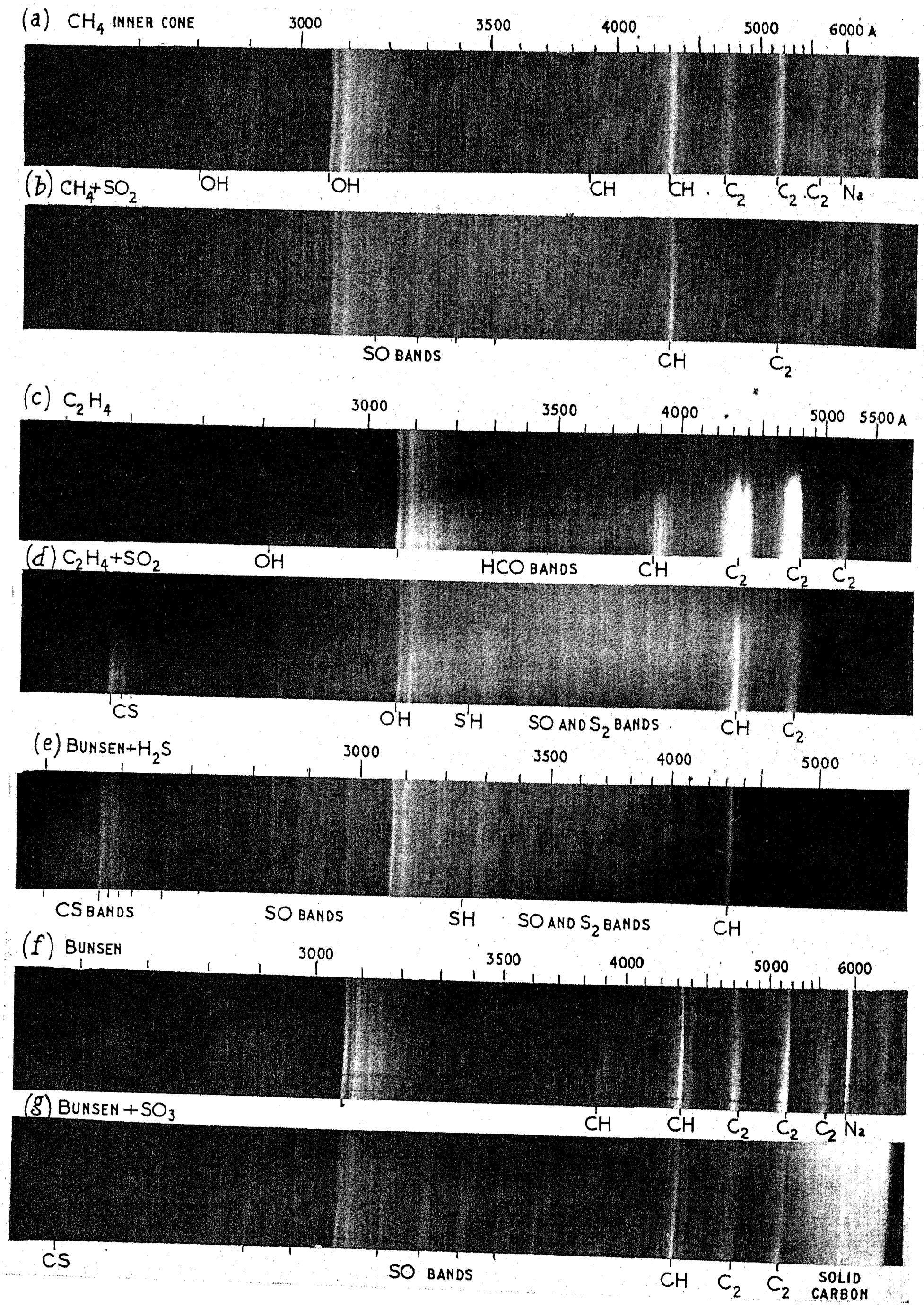


(*g*) CO FLAME



(*h*) CO + SO_2




(a) CH$_4$ INNER CONE
3000
3500
4000
5000
6000 A
(b) CH$_4$+SO$_2$
OH
OH
CH
CH
C$_2$
C$_2$
C$_2$
Na
SO BANDS
CH
C$_2$
(c) C$_2$H$_4$
3000
3500
4000
5000
5500 A
(d) C$_2$H$_4$+SO$_2$
OH
HCO BANDS
CH
C$_2$
C$_2$
C$_2$
CS
OH
SH
SO AND S$_2$ BANDS
CH
C$_2$
(e) BUNSEN+H$_2$S
3000
3500
4000
5000
CS BANDS
SO BANDS
SH
SO AND S$_2$ BANDS
CH
(f) BUNSEN
3000
3500
4000
5000
6000
(g) BUNSEN + SO$_3$
CH
CH
C$_2$
C$_2$
C$_2$
Na
CS
SO BANDS
CH
C$_2$
C$_2$
SOLID CARBON

Lenher, S. 1931 *J. Amer. Chem. Soc.* **53**, 3752.
Lenher, S. 1932 *J. Amer. Chem. Soc.* **54**, 1830.
Lepsoe, R. 1940 *Industr. Engng Chem.* **32**, 910.
Lewis, N. & White, J. U. 1939 *Phys. Rev.* **55**, 894.
Smith, E. C. W. 1940 *Proc. Roy. Soc.* A, **174**, 110.
Storch, H. 1934 *J. Amer. Chem. Soc.* **56**, 374.
Storch, H. 1935 *J. Amer. Chem. Soc.* **57**, 2598.
Taylor, H. S. & Jones, W. H. 1930 *J. Amer. Chem. Soc.* **52**, 1111.
Waterman, H. I. & van Vlodrop, C. 1939 *J. Soc. Chem. Ind., Lond.*, **58**, 109.
Whittingham, G. 1945 *Nature*, **156**, 207.

DESCRIPTION OF PLATES 8 AND 9

All spectrograms were taken on a small quartz spectrograph and except for strips *a, b, c, d* of plate 8, a wide slit was used throughout.

PLATE 8

	Exposure	Type of plate
a. Inner cone of Bunsen flame	60 min.	Ilford Golden Iso-Zenith
b. Inner cone of Bunsen flame containing 4 % SO_2	60 min.	,,
c. Luminous coal-gas flame (under-aerated Bunsen)	30 min.	,,
d. Luminous coal-gas flame (under-aerated Bunsen) + 4 % SO_2	4 hr.	,,
e. Inner cone of aerated H_2 flame containing 2 % SO_2 and burning at a quartz jet	20 min.	Zenith
f. Tip of outer cone of aerated $H_2 + 2$ % SO_2 flame. Fe arc comparison	15 min.	,,
g. Flame of CO burning at a quartz jet. Fe arc comparison	½ min.	,,
h. Flame of CO + 2 % SO_2 burning at a quartz jet	1½ min.	,,

PLATE 9

	Exposure	Type of plate
a. Inner cone of aerated methane flame	10 min.	H.P.3
b. Inner cone of aerated methane flame + 2 % SO_2	10 min.	,,
c. Aerated ethylene flame, inner and outer cones	3 min.	,,
d. Aerated ethylene flame, inner and outer cones + 2 % SO_2	3 min.	,,
e. Inner cone of Bunsen flame containing H_2S	4 min.	Zenith
f. Inner cone of Bunsen flame	3 min.	H.P.3
g. Inner cone of Bunsen flame + 0·1 % SO_3	3 min.	,,

On the simple group of order 25920

By J. A. Todd

(*Communicated by H. F. Baker, F.R.S.—Received* 4 *June* 1946)

Explicit expressions for the operations of the simple group of order 25920 as products of five or fewer operations of period two are obtained, leading to a determination of the conjugate sets of operations contained in the group. The results are applied to the determination of the nature of certain important subgroups, and to some questions relating to the generation of the group by two operations.

Introduction

The simple group G of order 25920 has formed the subject of a large number of investigations. It was first considered by Jordan (1870) as the group of even substitutions of the 45 tritangent planes of a general cubic surface. Jordan proved that the group was simple, and identified it with the group arising in the problem of trisection of the periods of quadruply periodic hyperelliptic functions. From this latter point of view the group was studied at length by Burkhardt, who obtained various representations of the group by linear transformations. In particular, he showed (Burkhardt 1890) that the group is simply isomorphic with a collineation group in space of four dimensions leaving a certain quartic primal invariant, and obtained a set of four particular operations which generate the group. The primal in question possesses 45 nodes, which are in $(1, 1)$ correspondence with the 45 tritangent planes of a general cubic surface, and form a remarkable configuration. The geometry of this configuration has been developed by Baker in his recent tract (Baker 1946). Baker has drawn attention to the importance of 45 particular operations of the group, which he has termed *projections*, each of which is a harmonic inversion having as centre one of the 45 nodes. He proves, among many other interesting results, that Burkhardt's four generators can all be expressed in terms of these projections, which accordingly generate the whole group.

The importance of these projections rests on the fact, proved below, that they form a complete set of conjugate operations in G. It will be proved in the present paper that every operation of the group can be expressed as a product of five or fewer of these projections, and shall express every operation of the group in such a form. It will be seen how this method of expressing the operations enables one to enumerate the various conjugate sets belonging to the group, confirming the results obtained (from a very different point of view) by Frame (1936); to determine readily the *periods* of the different types of operation; to enumerate the operations of certain important subgroups of G; and to verify results obtained by Coxeter, Brahana and Frame (referred to in § 8 below) concerning various simple ways in which G can be generated by a small number of operations, and to obtain other results of a similar nature.

1. DESCRIPTION OF A FUNDAMENTAL CONFIGURATION IN FOUR DIMENSIONS

This paper is concerned with the group G of collineations in four dimensions which leaves invariant a certain configuration of 45 points, and it is the purpose of the present section to describe this figure, and certain subsidiary figures contained in it, in sufficient detail to make the rest of the paper intelligible. Proofs of all the statements made here will be found in Baker's tract (Baker 1946), or may easily be deduced from Baker's results.

Take a coordinate system in S_4 in which a point is specified by the ratios of six numbers $x_1, x_2, \ldots, x_6$ whose sum is zero, and consider the $30+15$ points (ij, kl, mn), (ij) defined, respectively, by

$$x_i = x_j = 1, \quad x_k = x_l = \epsilon, \quad x_m = x_n = \epsilon^2,$$

and

$$x_i = 1, \quad x_j = -1, \quad x_k = x_l = x_m = x_n = 0,$$

where $(ijklmn)$ is a permutation of (123456) and ϵ is a primitive cube root of unity. It is easily seen that

$$(ij, kl, mn) = (ji, kl, mn) = (kl, mn, ij) = \text{etc.}, \quad (ij) = (ji).$$

These 45 points are the nodes of the quartic primal whose equation is

$$\Sigma\, x_i\, x_j\, x_k\, x_l = 0,$$

the summation extending over the 15 sets of four taken from among the suffixes $1, 2, \ldots, 6$; and it will be convenient to refer to them throughout the paper as *nodes*, although explicit reference to the primal will not be made in the sequel.

The group G of collineations, which keeps the set of nodes as a whole fixed, permutes the nodes *transitively* (that is, it contains operations which change any given node into any other), and it also permutes transitively the different configurations derived from these nodes which will now be described.

With each node is associated a definite prime, which will be called the *polar prime* of the node. The polar primes of the nodes (ij, kl, mn), (ij) have respective equations

$$(x_i+x_j)+\epsilon^2(x_k+x_l)+\epsilon(x_m+x_n) = 0, \quad x_i - x_j = 0.$$

The 45 primes so arising will be called, following Baker, *Jordan primes*.

With respect to any given one of the 45 nodes, A say, the remaining nodes fall into two sets, comprising 12 and 32 nodes respectively. The 12 nodes in the former of these two sets lie in the polar prime of A, and the line joining any one of these to A contains no further node. Such lines will be called *e-lines*. There are $\frac{1}{2}.45.12 = 270$ e-lines, each containing two of the nodes, 12 of these lines passing through each node. The 32 nodes which do not lie on e-lines through A lie in pairs on 16 lines through A. Such lines, which are called *κ-lines*, each contain three nodes, and there are $\frac{1}{3}.45.16 = 240$ of them, sixteen passing through each node. The polar primes of three nodes lying on a κ-line meet in a plane containing a second κ-line; two such κ-lines are symmetrically related and either is called the *polar κ-line* of the other. A pair of polar κ-lines consists, necessarily, of two skew lines, and the nine lines joining a node on one κ-line to a node on its polar are all e-lines.

It will be important in what follows to be able to recognize whether two nodes whose symbols are given are joined by an e-line or a κ-line, and in the latter case to determine the third node which lies on the κ-line. To this end the sets of three nodes which lie on κ-lines are listed; all such sets have one of the three forms

$$(ij), (jk), (ki);$$

$$(ij), (ik, jl, mn), (jk, il, mn);$$

$$(ij, kl, mn), (il, kn, mj), (in, kj, ml).$$

It will be noticed, in particular, that the 12 nodes lying in the polar prime of (ij) are these—other than (ij)—whose symbols are unaltered by the interchange of i and j, while pairs of nodes collinear with ij are such that their symbols are interchanged when i and j are interchanged.

There are planes, of four distinct types, each containing three or more nodes which are not collinear, as follows:

(i) There are 270 planes, which will be called *f-planes*, each containing three of the nodes and the three lines, which are e-lines, joining them in pairs. Each of these planes is the intersection of the polar primes of two other nodes, uniquely determined by the plane, whose join is an e-line. These nodes, together with the three that lie in the plane, form a symmetrical set of five nodes such that the polar prime of any one of them contains the other four; the join of any two of these nodes is an e-line and the plane containing any three is an f-plane. Such a set of five nodes is called a *Jordan pentahedron*, and there are 27 such, three of which contain any given node. In particular, the five nodes

$$(ij), \quad (kl), \quad (mn), \quad (ij, kl, mn), \quad (ij, mn, kl)$$

form a Jordan pentahedron, a fact which will be used in the sequel.

(ii) There are 720 planes, called *d-planes*, each of which joins a κ-line to a node lying on its polar κ-line. Each d-plane contains four nodes, three e-lines and one κ-line.

(iii) There are 540 planes, called *cross-planes* or *c-planes*, each containing six nodes, which are the vertices of a quadrilateral formed by κ-lines. Each c-plane contains six nodes, three e-lines (the diagonals of the quadrilateral) and four κ-lines.

(iv) There are 40 planes, called *Jacobian planes* or *j-planes*, each of which contains nine of the nodes, and twelve κ-lines. The nine nodes form the well-known configuration formed by the inflexions of a pencil of plane cubic curves; the join of any two of them is a κ-line and contains a third node, and the four κ-lines which pass through any one of the nodes form an equianharmonic pencil. The 12 κ-lines can be divided into four sets of three, each set forming a triangle whose sides contain all the nine nodes.

There are likewise primes, of four types, each containing four or more nodes which are not coplanar, as follows:

(i) It can easily be shown that each cross-plane lies in a definite Jordan prime, and that the polar κ-lines of the four κ-lines which lie in the plane pass through the pole of the prime (i.e. the node whose polar prime is the Jordan prime in question). The 540 primes, called *x-primes*, which join the cross-planes to the poles of the Jordan primes in which they lie, each contain seven nodes, nine e-lines, four κ-lines, three f-planes, four d-planes and one c-plane.

(ii) There are 216 primes, called *n-primes*, each containing ten nodes and ten κ-lines which are the vertices and edges of a pentahedron of cross-planes. Each n-prime contains 10 nodes, 15 e-lines, 10 κ-lines, 10 d-planes and five c-planes. The prime containing the ten nodes

$$(12),\quad (13),\quad (14),\quad (15),\quad (23),\quad (24),\quad (25),\quad (34),\quad (35),\quad (45)$$

is an example.

(iii) The 45 Jordan primes already mentioned each contain 12 nodes, which fall into three sets of four, forming a triad of desmic tetrahedra, any two of which are in perspective from each vertex of the third. Each of these tetrahedra, together with the pole of the prime, forms a Jordan pentahedron. Each Jordan prime contains 10 nodes, 18 e-lines, 16 κ-lines, 12 f-planes and 12 c-planes.

(iv) There are 40 *Steiner primes*, each of which contains four Jacobian planes which form a tetrahedron whose pairs of opposite edges are polar κ-lines; each edge of the tetrahedron thus contains three nodes. A Steiner prime contains 18 nodes, 27 e-lines, 42 κ-lines, 18 d-planes, 27 c-planes and four j-planes.

Elementary combinatorial considerations show that the lines, planes and primes just enumerated comprise all the linear spaces defined by subsets of the 45 nodes. Some further properties of the configuration are added which will be of use later.

There are four Steiner primes containing any given Jacobian plane. Each of the four triangles formed from three κ-lines in the plane which contain between them the nine nodes is a face of the tetrahedron of Jacobian planes in one of these primes. And any prime which contains a Jacobian plane and a node not lying in this is a Steiner prime. If Σ is a given Steiner prime, there are 12 other Steiner primes whose intersection with Σ is a Jacobian plane. The remaining 27 Steiner primes meet Σ in cross-planes, one such prime passing through each of the cross-planes in Σ. And the 27 Jordan primes which contain the cross-planes of Σ are all distinct, and are the polar primes of the 27 nodes which do not lie in Σ.

From the 27 Jordan pentahedra it is possible, in 72 ways, to select a set of six pentahedra whose vertices are all distinct. Each such set determines uniquely a second set whose pentahedra contain, collectively, the same 30 vertices. Thus there are 36 *double-sixes of pentahedra*. The 15 nodes which are not included among the vertices of a double-six are the vertices of a hexahedron of n-primes, and the (ten) pairs of opposite edges of the hexahedron are polar κ-lines. Such a set of nodes, of which there are 36, may be called a 15-set, a simple example being provided by the 15 nodes (ij).

2. THE SIMPLE GROUP G AND THE PROJECTIONS WHICH GENERATE IT

It is known (Coble 1906, p. 365) that it is possible to set up a (1, 1) correspondence between the 45 nodes of the figure described in § 1 and the 45 tritangent planes of a general cubic surface in such a way that the tritangent planes corresponding to two nodes A, B meet in a line of the cubic surface if AB is an e-line and have no line of the surface in common if AB is a κ-line. It follows that the group G^* of automorphisms of the 45 nodes (that is, the group of permutations of the nodes among themselves which preserve incidence relations) is simply isomorphic with the group of automorphisms of the tritangent planes. This group was studied by Jordan, who showed (Jordan 1870) that it contains a self-conjugate subgroup G which is simple and of order 25920; this group consists, in fact, of the group of automorphisms which effect an even permutation of the 45 tritangent planes (Dickson 1901, p. 306). I have shown elsewhere (Todd 1936, p. 173) that every operation of G^* corresponds to a collineation or anticollineation in S_4 which transforms the set of nodes into itself. The subgroup formed by the collineations belonging to G^* is of index 2, and it is clearly self-conjugate. Since G^* is known to contain only one such subgroup it follows that the group of collineations, of order 25920, is simply isomorphic with G (and will in future be denoted by G).

Baker has shown in his tract (Baker 1946, § 7) that the configuration is left unaltered by the harmonic inversion with respect to any node A and its polar prime. This harmonic inversion, which will be denoted by $p(A)$ and called the *projection* from the node A, is an operation of G which leaves unaltered the node A and the 12 nodes in the polar prime of A, and which interchanges in pairs the two nodes other than A on each of the 16 κ-lines which pass through A. Thus $p(A)$ effects an even permutation of the 45 nodes.

Suppose, now, that B, C, D are further nodes such that B and C lie on a κ-line through A, and D lies in the polar prime of A. Let T be any operation of G and let A', B', C', D' be the nodes into which A, B, C, D are respectively transformed by T. Then, since T is an automorphism of the configuration, $A'B'C'$ is a κ-line and D' lies in the polar prime of A'. Let $T' = Tp(A)\,T^{-1}$ be the transform of $p(A)$ by T. Then, since $p(A)$ leaves A and D unaltered and interchanges B and C,

$$\begin{aligned} T'[A'] &= Tp(A)\,T^{-1}[A'] = Tp(A)\,[A] = T[A] = A', \\ T'[B'] &= Tp(A)\,T^{-1}[B'] = Tp(A)\,[B] = T[C] = C', \\ T'[C'] &= Tp(A)\,T^{-1}[C'] = Tp(A)\,[C] = T[B] = B', \\ T'[D'] &= Tp(A)\,T^{-1}[D'] = Tp(A)\,[D] = T[D] = D', \end{aligned}$$

so that T' keeps fixed A' and every node in its polar prime, and interchanges the two nodes, other than A', on each κ-line through A'. Thus T' is the projection $p(A')$, for distinct operations of G effect distinct permutations of the nodes. It has thus been proved that

$$Tp(A)\,T^{-1} = p(A'), \quad \text{where} \quad A' = T[A].$$

Thus any operation of G which is conjugate to a projection is itself a projection. And, since G is transitive, there exist operations of G which transform A into any one of the other nodes. Thus all the 45 projections are conjugate. Hence *the* 45 *projections form a complete conjugate set in* G. The subgroup generated by them is thus self-conjugate in G, and hence, as G is simple, coincides with G itself. Thus Baker's statement that G is generated by projections is verified; and since each projection effects an even permutation of the nodes the same will be true for each operation of G. Observing, further, that the anticollineation $x_i' = \bar{x}_i$ keeps fixed the 15 nodes (ij) and interchanges the 15 pairs of nodes such as (ij, kl, mn), (ij, mn, kl), thus effecting an odd permutation of the nodes, it is seen at once that *every* anticollineation in G^* (which is the product of the one considered with a collineation of G) effects an odd permutation of the nodes, so that every even automorphism is determined by a collineation.

The fundamental relations connecting the various projections are given by Baker (1946, §7) and may be expressed in the following form (in which the identical operation of G is denoted by 1):

$$\left.\begin{aligned} [p(A)]^2 &= 1 \quad \text{for each node } A,\\ p(A)p(B) &= p(B)p(A) \quad \text{if } AB \text{ is an } e\text{-line},\\ p(A)p(B) = p(B)p(C) &= p(C)p(A) \quad \text{if } ABC \text{ is a } \kappa\text{-line},\\ p(A)p(B)p(C)p(D)p(E) &= 1 \quad \text{if } ABCDE \text{ is a Jordan pentahedron.} \end{aligned}\right\} \qquad (1)$$

I propose to show how, by means of these relations, every operation of G can be expressed as a product of five or fewer projections, to classify these operations and obtain their periods, and to determine the various conjugate sets contained in G. It will be seen that there are fifteen types of operation, the operations of any one type all being conjugate in the group of automorphisms G^*, and that in the case of five of these types the operations form two conjugate sets in G. For purposes of reference, these types will be denoted by Roman numerals, and I shall commence with two types whose existence is already clear, namely,

I. The identical operation.

II. The 45 projections, each of period 2, which form a conjugate set.

In determining the arrangement of the operations of G in conjugate sets frequent use will be made of the principle that if a set of nodes $A_1, \ldots, A_n$ is transformed by an operation T of G into a set $A_1', \ldots, A_n'$, then the operations $p(A_1)\ldots p(A_n)$ and $p(A_1')\ldots p(A_n')$, products of the projections from these sets of nodes, are conjugate, a result which is immediately evident if one observes that

$$\begin{aligned} Tp(A_1)\ldots p(A_n)\,T^{-1} &= Tp(A_1)\,T^{-1}\ldots Tp(A_n)\,T^{-1}\\ &= p(A_1')\ldots p(A_n'), \end{aligned}$$

by the result proved above. Two sets of nodes $A_1, \ldots, A_n$ and $A_1', \ldots, A_n'$ related in this way will be called equivalent with respect to G, and the aggregate of operations conjugate to a product of projections $p(A_1)\ldots p(A_n)$ consists of the products $p(A_1')\ldots p(A_n')$ of projections from all sets of nodes equivalent to $A_1, \ldots, A_n$.

In the course of subsequent work it will frequently be necessary to transform products of projections, usually with the aim of simplifying the product under discussion or of reducing it to some standard type. The transformations are based entirely on equations (1). But one particular type of transformation is of such frequent occurrence that it is described explicitly here. From (1) it follows that any product $p(A)p(B)$ of two projections can be expressed in each of the forms $p(B)\,p(A')$ or $p(B')\,p(A)$, where, if AB is an e-line, $A' = A$ and $B' = B$, while if AB is a κ-line, $A' = B' = C$, where C is the remaining node of the κ-line. It follows that in any product of projections any given factor $p(A)$ may be moved one or more places to the right (or left) without altering the number of projections in the product and without altering any factor which is not passed over by $p(A)$. More generally, two or more factors may be moved about in the same way (with consequent changes only in the factors passed over) so long as they retain their original order. Any transformation of this type will be referred to as an *elementary transformation* of the product, and it will be assumed that such transformations will be recognized as such by the reader after a few examples have been given.† The other transformations involve either the cancelling of two consecutive factors $p(A), p(A)$, which calls for no comment, or the replacement of the product of projections from a set of nodes belonging to a Jordan pentahedron by the product of projections from the remaining nodes of the pentahedron. When this is likely to cause any difficulty to the reader a footnote has been added by way of explanation.

By way of example two elementary results are proved. Calling a product of projections *reducible* if it can be expressed as a product involving fewer factors the following theorems are proved:

(i) *Any product which contains a repeated factor is reducible.*

(ii) *Any product which contains three distinct factors $p(A)$, $p(B)$, $p(C)$—not necessarily consecutive—is reducible if either ABC is a κ-line or ABC is an f-plane.*

To prove (i) it is sufficient to observe that by elementary transformations the product can be transformed into one in which the repeated factor $p(A)$ occurs as two consecutive factors; and then, since $p(A)$ has period 2, these two factors can be removed altogether. Similarly, in case (ii), elementary transformations will reduce the product to one in which $p(A), p(B), p(C)$ are consecutive factors, suppose in this order. Then, if ABC is a κ-line, $p(A)p(B)p(C) = p(B)p(C)p(C) = p(B)$, and if ABC belong to a Jordan pentahedron $ABCDE$, then, by the last equation of (1), $p(A)p(B)p(C) = [p(D)p(E)]^{-1} = p(E)p(D)$, and in each case the number of factors has been reduced.

3. PRODUCTS OF TWO PROJECTIONS

In considering the products of projections from two distinct nodes A, B there are two cases to be distinguished, according as AB is an e-line or a κ-line.

If AB is an e-line, equations (1) show that $p(A)$ and $p(B)$ commute, so that from each e-line a single product $p(A)p(B)$ of this type arises, and this operation is of period 2. Thus $p(A)$ and $p(B)$ generate an Abelian group of order 4, direct product of the cyclic groups generated by $p(A), p(B)$. And it is easy to see that $p(A)p(B)$ is an axial homology having for invariant spaces the line AB and the opposite plane CDE of the Jordan pentahedron of which AB is an edge. Thus $p(A)p(B)$ leaves the nodes A, B, C, D, E unchanged, and only these nodes. Whence it is easily seen that the operations of this type arising from different e-lines are distinct. And since G permutes the e-lines transitively, every such operation is conjugate to $p(A)p(B)$. Hence, since there are 270 e-lines, G contains

---

† The list of types of κ-line in § 1 will be found useful here.

III. *A conjugate set of* 270 *operations of period* 2, *products of pairs of projections from nodes lying on an e-line.*

If AB is a κ-line, and if C is the third node lying on AB, then, from (1),

$$p(A)p(B) = p(B)p(C) = p(C)p(A),$$

and so
$$[p(A)p(B)]^3 = p(A)p(B).p(B)p(C).p(C)p(A) = 1,$$

so that $p(A)p(B)$ is of period 3. Its square (or inverse) is $p(B)p(A)$, and there are just two products of pairs of projections from nodes on ABC which are distinct; the projections $p(A), p(B), p(C)$ generate a group of order 6, which, as the reader will easily verify, is simply isomorphic with the symmetric group of degree 3 (and in fact permutes the nodes A, B, C according to the operations of this group). The operation $p(A)p(B)$ keeps fixed the three nodes of the polar κ-line of ABC, and no other node. Hence the operations arising from distinct κ-lines are distinct. Since G is transitive on the κ-lines, any operation of this type is conjugate to one of the two operations $p(A)p(B), p(B)p(A)$. And these are clearly conjugate, since each is transformed into the other by $p(A), p(B)$ or $p(C)$. Thus, since there are 240 κ-lines, G contains

IV. *A conjugate set of* 480 *operations of period* 3, *each the product of projections from two nodes of a κ-line.*

4. PRODUCTS OF PROJECTIONS FROM COPLANAR SETS OF NODES

4.1. Next consider the products of projections from sets of coplanar nodes. In examining such products one may, by the theorems proved at the end of § 2, ignore products containing repeated factors or products involving projections from three collinear nodes. There are four cases to consider, since there are four types of plane containing three or more non-collinear nodes. In the case of the f-plane or the d-plane the results are so easily seen that they are stated without proof, their verification being left to the reader.

No new type of operation arises from an f-plane, since the product of the three projections from the nodes lying in it is reducible (§ 2). The group generated by the projections is an Abelian group of order 8, direct product of three cyclic subgroups of order 2.

In the case of a d-plane, joining a κ-line ABC to a node P on its polar κ-line, it is easily seen that the projections from A, B, C, P generate a group of order 12, the direct product of the group of order 6 generated by $p(A), p(B), p(C)$ and the cyclic group generated by $p(P)$, the projection $p(P)$ commuting with every operation of this group. There are just two operations of a new type, namely, $p(A)\,p(B)\,p(P)$ and its inverse $p(B)\,p(A)\,p(P)$, and these operations are conjugate, being transformed into each other by $p(A), p(B)$ or $p(C)$. The operation $p(A)\,p(B)\,p(P)$ is of period 6, its square being $p(B)\,p(A)$ and its cube $p(P)$, so that different d-planes give different operations. Thus, since there are 720 d-planes, G contains

V. *A conjugate set of* 1440 *operations of period* 6, *products of three projections, whose squares are of type IV and whose cubes are of type II.*

4·2. Next consider the operations derived from sets of nodes in a c-plane. Such a plane, as was seen in § 1, contains six nodes forming the vertices of a complete quadrilateral, whose sides are κ-lines and whose diagonals are e-lines. Denote these nodes by A, B, C, A', B', C' in such a way that the sides of the quadrilateral are $A'BC, AB'C, ABC', A'B'C'$ and the diagonals are AA', BB', CC'.

It can be shown, in a variety of ways, that the projections from the six nodes in a c-plane generate a group of order 24, simply isomorphic with the symmetric group of degree 4. Thus, for instance, one such plane contains the nodes (12), (13), (14), (23), (24), (34); and, since the effect of $p(ij)$ on the nodes is represented by the interchange of i and j in the symbols of the nodes, the operations of the group generated by the six projections effect the 24 permutations of 1, 2, 3, 4 in the symbols. Alternatively, since the projection from any node C' on a κ-line $A'B'C'$ is expressible in the form $p(A')\,p(B')\,p(A')$, the projections from the six nodes A, B, C, A', B', C' can all be expressed in terms of $p(A), p(B), p(A')$; and these projections satisfy the relations

$$[p(A)]^2 = [p(B)]^2 = [p(A')]^2 = [p(A)\,p(B)]^3 = [p(B)\,p(A')]^3 = [p(A)\,p(A')]^2 = 1,$$

which are defining relations for the symmetric group of degree 4 (Moore 1897; or Burnside 1911, p. 464).

Of the 24 operations of this group all but six belong to types already considered, for besides identity and six projections it is seen that there are three operations of type III arising from the three diagonals of the quadrilateral and eight operations of type IV, two arising from each of the four κ-lines. The remaining six operations may thus be identified with the six operations of period 4 in the symmetric group. Each of these operations can in fact be expressed, in a variety of ways, as the product of three projections. I shall content myself by showing, directly from the geometry of the figure, that they are in fact obtained by permuting A, B, C in all possible ways in the product $p(A)\,p(B)\,p(C)$. It is an easy exercise in the use of elementary transformations to express such a product in terms of any other set of three projections from nodes whose joins are three of the four κ-lines in the plane. Observe, first, that $p(A)\,p(B)\,p(C)$ permutes the six nodes according to the permutation $(ACA'C')(BB')$, which shows that the six operations of this type are distinct. Further, by elementary transformations,

$$\begin{aligned}[p(A)p(B)p(C)]^2 &= p(A)p(B)p(C).p(A)p(B)p(C)\\ &= p(A)p(B)p(C).p(C)p(B')p(A') = p(A)p(B)p(B')p(A')\\ &= p(B)p(C')p(C')p(B') = p(B)p(B'),\end{aligned}$$

an operation of type III and period 2. Thus $p(A)p(B)p(C)$ is of period 4. And the transforms of $p(A)p(B)p(C)$ by $p(A)$ and $p(A')$ are easily seen to be $p(B)p(C)p(A)$ and $p(A)p(C)p(B)$. From this it follows that the six operations in question are all conjugate in G. Hence, as G permutes the cross-planes transitively, the operations of this type form a single conjugate set in G. Now the operations arising from different cross-planes are necessarily distinct. For suppose, if possible, that $A_1 B_1 \ldots C_1'$ is a second cross-plane, the arrangement of the nodes on κ-lines being the same as in the plane $AB \ldots C'$ (with the suffix 1 of course), and that

$$p(A_1)p(B_1)p(C_1) = p(A)p(B)p(C).$$

Then the squares of these operations, namely, $p(B_1)p(B_1')$ and $p(B)p(B')$, are equal. Hence the e-lines B_1B_1' and BB' are the same, and B_1, B_1' coincide in some order with B and B'. Since

$$p(A_1)p(B_1)p(C_1) = p(A_1)p(C_1)p(A_1') = p(C_1)p(B_1')p(A_1'),$$

it may be assumed, without loss of generality, that $B_1 = B, B_1' = B'$. But then, since

$$p(A_1)p(B_1)p(C_1) = p(A_1)p(A_1')p(B_1), \quad p(A)p(B)p(C) = p(A)p(A')p(B),$$

it is seen that $p(A_1)p(A_1') = p(A)p(A')$. Thus, A_1, A_1' coincide in some order with A, A' and the two planes must be identical. Hence, since each of the 540 c-planes gives rise to six operations of this kind, G contains

VI. *A conjugate set of* 3240 *operations of period* 4, *products of projections from three nodes of a cross-plane, whose squares are of type III.*

4·3. The situation arising in the case of a j-plane is not quite so simple. The nine nodes belonging to a j-plane lie by threes on 12 κ-lines, and the join of any two of these nodes passes through a third. It is familiar that the arrangement of such a set of nine points in triads of collinear points may be represented by the scheme

$$\begin{array}{ccc} 1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{array}$$

in which 1, 2, ..., 9 denote the nine points, and the triads of collinear points are given by the rows, columns, diagonals and broken diagonals in the array. The nine nodes of a Jacobian plane will be denoted by 1, 2, ..., 9 in the present section, and from the array just given it is seen that the 12 κ-lines are

$$123,\quad 456,\quad 789,\quad 147,\quad 258,\quad 369,\quad 159,\quad 267,\quad 348,\quad 168,\quad 249,\quad 357.$$

It will also be convenient to denote the projection from the node i by $\{i\}$, the product $\{i\}\{j\}$ by $\{ij\}$ and so on.

Any irreducible product of three projections from nodes in the plane has the form $\{ijk\}$, where i, j, k are three distinct nodes not lying on a κ-line. There are $9.8.6 = 432$ such products, and it will be proved that these give rise to 18 distinct operations of G, each product being expressible in 24 equivalent forms. In fact, taking a typical product $\{124\}$ and using the list of κ-lines just given, one can verify without difficulty that the 24 products

$$\begin{array}{cccccccc}\{124\}, & \{234\}, & \{314\}, & \{479\}, & \{512\}, & \{719\}, & \{837\}, & \{928\},\\ \{149\}, & \{248\}, & \{347\}, & \{498\}, & \{523\}, & \{795\}, & \{875\}, & \{985\},\\ \{192\}, & \{283\}, & \{371\}, & \{487\}, & \{531\}, & \{751\}, & \{853\}, & \{952\},\end{array}$$

are all equal. Moreover, $\{124\}$ and its inverse $\{421\}$ keep fixed the node 6 and interchange in pairs the nodes 1, 8; 2, 7; 3, 9; 4, 5 which lie on κ-lines through 6. Hence the 432 products give rise to nine pairs of inverse operations such as $\{124\}$ and $\{421\}$, each pair keeping fixed a definite one of the nine nodes and interchanging in pairs the nodes lying on κ-lines through this. It remains to prove that $\{124\}$ and $\{421\}$ are distinct. Now

$$\{124\}^3 = \{124\}\{479\}\{928\} = \{12728\} = \{168\} = \{688\} = \{6\},$$

so that $\{124\}^3$ is the projection $\{6\}$. And if $\{124\} = \{421\}$ then $\{124\}^2 = 1$, so that $\{124\} = \{124\}^3 = \{6\}$ and hence $\{12\} = \{64\}$, in contradiction with the result proved in § 3 that products of projections from two nodes of different κ-lines are distinct. Thus there are precisely 18 operations such as $\{ijk\}$ which are distinct, and each of these is an operation of period 6.

Next observe that in the 24 different expressions of $\{124\}$ as a product of projections $\{ijk\}$ the last factor $\{k\}$ takes every value except $\{6\}$. Thus the only possible irreducible product of four projections of the form $\{124l\}$ is $\{1246\}$. And since any set of five or more of the nine nodes includes at least three which are collinear, no product of more than four projections from the nodes in the plane can be irreducible. Thus there remain for consideration 18 operations $\{ijkl\}$, where $\{ijk\}$ is one of the 18 operations of period 6 just considered and l is the node left invariant by $\{ijk\}$. It will be proved that in fact all these operations are equal either to $\{1246\}$ or $\{4216\}$. Considerations of symmetry show that it suffices to consider a single pair of operations $\{ijkl\}, \{kjil\}$, where $l \neq 6$.

Observe, first, that since $\{124\}^3 = \{6\}$ and $\{421\} = \{124\}^{-1}$, then

$$\{1246\} = \{124\}^4 = \{421\}^2, \quad \{4216\} = \{124\}^2,$$

so that $\{1246\}$ and $\{4216\}$ are inverse operations, each of period 3. It is thus merely necessary to show that the squares of another typical pair of inverse operations $\{ijk\}, \{kji\}$ coincide with $\{1246\}$ and $\{4216\}$. Taking, for instance, the operations $\{125\}$, $\{521\}$, which leave the node 4 unchanged, it is seen that

$$\begin{aligned}\{125\}^2 &= \{125125\} = \{125598\} = \{1298\} = \{1428\} = \{4728\}\\ &= \{4268\} = \{4216\},\\ \{521\}^2 &= \{125\}^{-2} = \{4216\}^{-1} = \{1246\}.\end{aligned}$$

There are thus just two operations of this type, and the square of any one of the 18 operations of period 6 coincides with one of them. Hence the projections from the nodes of a Jacobian plane generate a group of order 54, comprising identity, nine projections, 24 operations of type IV and the 18 operations of period 6 and the two operations of period 3 just obtained.

It may be remarked, in passing, that the operation S of period 3 which is one of Burkhardt's generating operations, is of the last type discussed, and several expressions for it as a product of four projections are given on p. 77 of Baker's tract.

Each of the operations $\{1246\}$, $\{4216\}$ leaves invariant every node in the Jacobian plane. It can leave no other node unaltered. For considerations of incidence show that any operation of G which keeps fixed all the nodes in a j-plane and a node P not lying in this plane keeps fixed each node in the Steiner prime joining the j-plane to P, so that it keeps fixed each c-plane in this prime and hence the remaining 27 nodes, whose polar primes intersect the Steiner prime in these c-planes; such an operation of G must therefore be the identity. It follows that the two operations of period 3 derived from each of the 40 j-planes give 80 distinct operations. And hence the operations of period 6 derived from different j-planes, whose squares are operations of period 3 of the type just mentioned, must also be distinct.

It will now be proved that the operations of each of the two types mentioned form two conjugate sets in G. For this observe, first, that as G is transitive on the j-planes every operation (of either type) is conjugate to an operation of the same type associated with any given j-plane. And since the group generated by the projections $\{i\}$ permutes the nine nodes i transitively, any operation of period 6 in the plane is conjugate to one or other of the operations $\{124\}$, $\{421\}$ which leave the node 6 unchanged. Now, if there exists an operation T of G such that $T\{124\}T^{-1} = \{421\}$, then, also, $T\{124\}^2 T^{-1} = [T\{124\}T^{-1}]^2 = \{421\}^2$. Thus, if $\{124\}$ and $\{421\}$ are conjugate in G, so also are $\{1246\}$ and $\{4216\}$. It will now be shown that this is not the case.

In fact, if there exists an operation T in G such that $T\{1246\}T^{-1} = \{4216\}$, then the nine nodes $i' = T(i)$ lie in a Jacobian plane, and the set of four nodes $1'2'4'6'$ is projective with 1246 and such that $\{1'2'4'6'\} = \{4216\}$. Since distinct Jacobian planes give rise to distinct operations it follows that the plane of the nodes i' coincides with that of the nodes i. Thus there is a projectivity in the Jacobian plane which keeps the set of nine nodes unaltered and transforms i into i'. Now the group of projectivities which leave the set of nine nodes unaltered is known to be of order 216, and the group of automorphisms of the set of nodes is of order 432, and consists of

216 collineations and 216 anticollineations. There are precisely 432 ordered sets of nodes ijk which are not collinear, and there exists a unique automorphism of the nine nodes which transforms any one of these sets into any other. Each such triad of nodes determines uniquely a fourth node l such that $\{ijkl\}$ leaves each node unaltered. Thus, if $\{ijkl\}$ and $\{i'j'k'l'\}$ are two expressions of this type, there is a unique automorphism which transforms i, j, k, l respectively into i', j', k', l'. It is not difficult to see that any two operations of this type which can be transformed into one another by elementary transformations must be such that the corresponding automorphism is a collineation. Hence, if $\{1246\}$ and $\{4216\}$ are conjugate, there is a projectivity T_1 leaving the set of nodes invariant as a whole which transforms $1', 2', 4', 6'$ into $4, 2, 1, 6$ respectively. Hence, if T_2 is the projectivity induced among the nine nodes by T, T_1T_2 is a projectivity of the set of nodes in which $1, 2, 4, 6$ correspond respectively to $4, 2, 1, 6$. In this projectivity the node 3 on the line 12 must correspond to the node 9 on the line 42. Hence the pencils $6(1243)$ and $6(4219) = 6(4213)$ are projective and consequently harmonic, whereas they are known to be equiunharmonic. Hence $\{1246\}$ and $\{4216\}$ are not conjugate in G (although they are conjugate in G^*, being transformed into each other by an anticollineation). So, from the Jacobian planes, one obtains

VII. 80 *operations of period* 3, *products of four projections, forming two conjugate sets of* 40 *operations each,*

and

VIII. 720 *operations of period* 6, *products of three projections, forming two conjugate sets of* 360 *operations each, whose squares are of type* VII *and whose cubes are of type* II.

5. PRODUCTS OF PROJECTIONS FROM SETS OF NODES IN A PRIME

5·1. Now consider the products of projections from sets of nodes lying in a prime, and examine, in turn the four types of prime described in § 1. The possibilities presented by an x-prime are soon disposed of. For an x-prime is the join of a cross-plane to the pole of the Jordan prime which contains it. If, as in § 4·2, the six nodes of the cross-plane are denoted by A, B, C, A', B', C' and the pole of the Jordan prime containing it by O, the projection $p(O)$ is easily seen to commute with the projections from the other six nodes. Hence the group generated by the projections from the nodes of the prime is of order 48, and is the direct product of the symmetric group of degree 4 generated by the projections from the nodes in the cross-plane and the cyclic group of order 2 generated by $p(O)$. And the only operations of this group which do not clearly belong to types already obtained are the products of $p(O)$ with the six operations of period 4 (type VI) belonging to the cross-plane. But, in the notation of § 4·2,

$$p(A)p(B)p(C)p(O) = p(B)p(C')p(C)p(O),$$

and this product is reducible since the plane OCC' is an f-plane. Thus no new type of operation arises. Incidentally, since G permutes the 540 x-primes transitively, the group of order 48 just mentioned is just the subgroup of G which keeps the

x-prime fixed. It is easily seen that the 48 operations of this group comprise, besides identity, 7 operations of type II, 12 of type III, 8 of type IV, 8 of type V, and 12 of type VI.

The operations derived from the nodes of an n-prime are likewise easily discussed. For a typical n-prime is that containing the ten nodes (ij) in which neither i nor j is 6, and since the effect of $p(ij)$ on the nodes is represented by the interchange of i and j in the symbols of these nodes it follows that the projections from the ten nodes (ij) generate a group of order 120, simply isomorphic with the symmetric group of degree 5. And, since there are 216 n-primes permuted transitively by the operations of G, this group is simply the subgroup of G which leaves the n-prime invariant. It is easy to see that the only operations of this subgroup which are not of a type already considered are the 24 operations of period 5 which permute the symbols 1, 2, 3, 4, 5 in some cyclic order. And since any cyclic permutation $(ijklm)$ is equivalent to the product of four transpositions—for example, to the product $(ml)\,(mk)\,(mj)\,(mi)$—it will be seen that each of these operations is expressible as the product of four projections, the product $p(ml)\,p(mk)\,p(mj)\,p(mi)$ being a typical one. These 24 operations are conjugate in the symmetric group of degree 5, and hence are conjugate in G, which contains this symmetric group as a subgroup. Hence, since G permutes the 216 n-primes transitively, all the operations of this type form a complete set of conjugates in G. Each of these operations permutes the ten nodes of the n-prime from which it arises in two cycles of 5, the five nodes in any cycle not all lying in a plane. But it is easily seen, from the incidence relations of § 1, that any two n-primes have at least one of the nodes in common. Whence it follows that no two operations arising from different n-primes can be identical. Thus G contains

IX. *A conjugate set of* 5184 *operations of period* 5, *each the product of four projections.*

The 120 operations which keep the n-prime fixed comprise identity, ten operations of type II, 15 of type III, 20 of type IV, 20 of type V, 30 of type VI, and 24 of type IX.

5·2. Consider next a Jordan prime Π, whose pole is denoted by O. Then Π contains 12 nodes forming the vertices of three desmic tetrahedra, any one of which, together with O, forms a Jordan pentahedron. These tetrahedra will be denoted by σ_i $(i = 1, 2, 3)$ and the vertices of σ_i by A_i, B_i, C_i, D_i; and symbols such as $P_i, Q_i, \ldots$ will be used to refer to unspecified vertices of σ_i. The line joining any vertex of σ_1 to any vertex of σ_2 contains a vertex of σ_3, and the notation may be chosen in such a way that the vertex of σ_3 is determined by the scheme

	A_2	B_2	C_2	D_2
A_1	A_3	B_3	C_3	D_3
B_1	B_3	A_3	D_3	C_3
C_1	C_3	D_3	A_3	B_3
D_1	D_3	C_3	B_3	A_3

Any product of projections from a set of non-coplanar nodes involes four factors at least, and so, if the nodes lie in Π, must contain at least two projections from nodes of the same tetrahedron. If it contains three such projections it is reducible. Thus any irreducible product of projections from non-coplanar nodes in Π can be reduced by elementary transformations to the form $p(P_i)p(Q_i)T$, where T is a product of two or more projections from nodes of the tetrahedra σ_j, σ_k distinct from σ_i. If T contains as factors two projections $p(R_j), p(S_k)$ from nodes of different tetrahedra, then since the κ-line R_jS_k contains a node belonging to σ_i one can transform $p(P_i)p(Q_i)T$ into a form involving three projections from nodes of σ_i, and this product is accordingly reducible. It follows at once that every product of five or more projections from nodes of Π is reducible, and that any irreducible product of four projections must be expressible in the form

$$p(P_i)p(Q_i)p(R_j)p(S_j),$$

where $P_i Q_i, R_j S_j$ are non-intersecting edges of distinct tetradedra σ_i, σ_j. Since the edges of the tetrahedra are e-lines the projections from P_i, Q_i (or R_j, S_j) commute. Thus there are essentially $18 \cdot 8 = 144$ products of this type to consider. It will be shown that these determine 12 distinct operations.

Consider, for instance, the typical product $p(A_1)p(B_1)p(A_2)p(C_2)$. Then, by elementary transformation,

$$p(A_1)p(B_1)p(A_2)p(C_2) = p(A_2)p(A_3)p(B_3)p(C_2),$$

and the latter product can be transformed, in two ways, to give

$$p(A_2)p(C_2)p(C_1)p(D_1) \quad \text{and} \quad p(A_3)p(B_3)p(B_2)p(C_2).$$

By means of transformations of this type it is seen that in any product

$$p(P_i)p(Q_i)p(R_j)p(S_j)$$

of the type considered the two tetrahedra σ_i, σ_j to which P_i, $Q_i, R_j S_j$ belong may be assigned in advance. Moreover,

$$\begin{aligned} p(A_1)p(B_1)p(A_2)p(C_2) &= p(A_1)p(B_1)p(O)p(O)p(A_2)p(C_2) \\ &= p(C_1)p(D_1)p(B_2)p(D_2), \end{aligned}$$

since O forms a Jordan pentahedron with each of the tetrahedra σ_i, so that the same operation arises if each of the edges $P_i Q_i, R_j S_j$ is replaced by the opposite edge of the tetrahedron to which it belongs. There are thus at most 12 distinct operations of this type, all expressible in the form $p(A_1)p(P_1)p(R_2)p(S_2)$, where P_1 is a vertex of σ_1 distinct from A_1 and R_2S_2 is one of the four edges of σ_2 which do not meet A_1P_1. These operations are easily seen to be distinct, for $p(P_1)p(R_2)p(S_2)$ is an operation of type VI, associated with the cross-plane $P_1R_2S_2$ (which does not contain A_1), interchanging the two nodes of this cross-plane which belong to σ_1, and two distinct (inverse) operations of this kind arise from each of the six cross-planes in Π which do not contain A_1. Further, since

$$\begin{aligned}[p(A_1)p(B_1)p(A_2)p(C_2)]^2 &= p(A_1)p(B_1)p(A_2)p(C_2).p(A_2)p(C_2)p(C_1)p(D_1)\\ &= p(A_1)p(B_1)p(C_1)p(D_1) = p(O),\end{aligned}$$

each of these operations is of period 4, and operations arising from different Jordan primes have different squares and so are distinct. By using identities such as

$$\begin{aligned}p(A_2).p(A_1)p(B_1)p(A_2)p(C_2).p(A_2) &= p(A_1)p(B_1)p(B_2)p(C_2),\\ p(B_1).p(A_1)p(B_1)p(A_2)p(C_2).p(B_1) &= p(A_1)p(D_1)p(A_2)p(C_2),\end{aligned}$$

it is easy to see that the 12 operations derived from Π are all conjugate. Hence, as G is transitive on the Jordan primes, the operations of this type in G form a single conjugate set, and hence G contains

X. *A conjugate set of* 540 *operations of period* 4, *products of four projections, whose squares are of type II.*

The group generated by the projections from the nodes in Π may now be determined. Since $p(O)$ belongs to the group it is necessary to consider not merely the operations of various types derived from nodes in Π but also the products of these with $p(O)$, when these products are irreducible. It is not difficult to verify that this group is of order 192, and that its operations consist of (i) identity, (ii) the projection $p(O)$, (iii) the 12 projections from nodes in Π, (iv) the 12 products of these with $p(O)$, which are of type III, (v) the 18 products of projections from two nodes on an edge of one of the tetrahedra in Π (type III), the product of such an operation by $p(O)$ being the corresponding operation derived from the opposite edge of the tetrahedron, (vi) 32 operations of type IV, two arising from each of the 16 κ-lines in Π, (vii) the 32 products of these with $p(O)$, which are of type V, (viii) 72 operations of type VI, six arising from each of the 12 c-planes in Π, the product of any one of which with $p(O)$ is another operation of the same set, (ix) the 12 operations of type X arising from Π, the product of any one of which with $p(O)$ is the inverse operation. This group has the cyclic group generated by $p(O)$ as a self-conjugate subgroup, the two operations of which leave fixed each node in Π, and the nodes of Π are permuted by the operations of the group of order 192 according to the 96 operations of the factor group of this cyclic group. The group of order 192 is, as will be seen presently, a subgroup of the group of order 576 consisting of all the operations of G which leave Π and its pole O invariant.

5·3. Next consider the products of projections from nodes lying in a Steiner prime Σ. Such a prime contains 18 nodes lying by threes on the edges of a tetrahedron whose faces are Jacobian planes and whose edges form three pairs of polar κ-lines. Baker has given, on p. 79 of his tract, explicit equations for the projections from the 18 nodes of one typical Steiner prime. It will be found convenient to denote the nodes of Σ by the symbols used by Baker to denote these projections. Accordingly, the nodes will be denoted by $X, Y, Z, P, Q, R, U, V, W, X', Y', Z', P', Q', R', U', V', W'$, the notation being such that $XYZ, X'Y'Z'$; $PQR, P'Q'R'$; $UVW, U'V'W'$ are the three pairs x, x'; p, p'; u, u' of opposite edges of the tetrahedron. Reference to Baker's tract shows that the arrangements of the nodes in the four Jacobian planes are given by the scheme

$$\begin{array}{llll} XYZ & XYZ & X'Y'Z' & X'Y'Z' \\ PRQ & P'R'Q' & PQR & P'Q'R' \\ WVU & W'V'U' & V'U'W' & VUW \end{array}$$

where, in each set of nine nodes, the arrangement on κ-lines is determined, as in §4·3, by the rule that three nodes on a κ-line occur either all in the same row or column or else each in a different row and column.

The examination of sets of nodes of Σ which give rise to irreducible products of projections may be carried out by methods similar to these used in the discussion of the Jordan prime, but this procedure is somewhat laborious and not particularly interesting, although it was the method by which the types of operation obtained in this section were originally derived. I shall proceed here in a different manner, first examining certain products which prove to yield new types of operation, and then showing that all irreducible types have been discovered by proving that all the operations of the group generated by the projections from nodes of Σ (which is necessarily a subgroup of the group of operations of G which keep Σ fixed, and proves in fact to coincide with this group) are accounted for by operations of types already determined.

The first type of product considered is a product of four projections of the form T_1T_2, where T_1 is the product of projections from two nodes on an edge of the tetrahedron and T_2 is the product of projections from two nodes on the opposite edge. Since opposite edges of the tetrahedron are polar κ-lines T_1 and T_2 commute. Hence, as each κ-line gives rise to two distinct products of two projections, there are 12 products of this type to consider (not regarding T_1T_2 and T_2T_1 as distinct). As a matter of fact these 12 products give rise to six distinct operations, each being expressible in two essentially different ways. In fact, as is stated by Baker, there are identities of the form

$$p(Y)p(Z)p(Y')p(Z') = p(V)p(W)p(V')p(W'),$$

which may be proved by noticing that

$$\begin{aligned} p(Y)p(Z)p(Y')p(Z') &= p(Y)p(Z)p(P).p(P)p(Y')p(Z') \\ &= p(P)p(U)p(V).p(W')p(U')p(P) \\ &= p(U)p(V)p(R).p(R)p(W')p(U') \\ &= p(U)p(V)p(W')p(U') = p(V)p(W)p(V')p(W'). \end{aligned}$$

The typical operation $p(Y)p(Z)p(Y')p(Z')$ is easily seen to keep fixed each node on p, p' and to permute cyclically the nodes on each of the other edges of the tetrahedron. Whence it is easily seen that there are six distinct operations of this type associated with Σ, and that these can be transformed into one another by means of projections from nodes lying in Σ, so that they are all conjugate. Moreover, no κ-line of Σ other than the edges of the tetrahedron, and no c-plane lying in Σ (which contains one node lying on each edge of the tetrahedron) is left unaltered by this operation. Whence different Steiner primes must give rise to distinct operations. Since G permutes the Steiner primes transitively, all these operations are conjugate in G. Since T_1 and T_2 are commuting operations of period 3, T_1T_2 has period 3. Hence G contains

XI. *A conjugate set of* 240 *operations of period* 3, *products of four projections, each expressible in two ways as a product of two operations of type III arising from a pair of polar κ-lines.*

The next type of product to be considered is of the form $T_1 . p(A)$, where T_1 is an operation of type VIII, product of three projections from nodes in a Jacobian plane in Σ leaving a node B of this plane fixed, and A is a node lying on the edge of the tetrahedron opposite to that which contains B. There are 18 operations such as T_1 associated with any one of the Jacobian planes, and three positions possible for A corresponding to any operation T. Thus each Jacobian plane gives rise to 54 products of this type. It will be proved that the same set of operations arises from each of the four Jacobian planes in Σ, and that the 54 products which arise are all distinct.

Consider the typical operation $T = p(X)p(Y)p(U)p(U')$, where $p(X)p(Y)p(U)$ is an operation of type VIII associated with the Jacobian plane xpu leaving the node V fixed, and U' lies on the polar κ-line of the edge u of the tetrahedron which contains V. Then

$$\begin{aligned} p(X)p(Y)p(U)p(U') &= p(X)p(Y)p(U')p(U), \\ p(X)p(Y)p(U)p(U') &= p(U')p(R')p(P')p(U) = p(R')p(P')p(W')p(U) \\ &= p(R')p(P')p(U)p(W'), \\ p(X)p(Y)p(U)p(U') &= p(U)p(R)p(P)p(U') = p(R)p(P)p(W)p(U') \\ &= p(R)p(P)p(U')p(W), \end{aligned}$$

and $$p(X)p(Y)p(U'), \quad p(R')p(P')p(U), \quad p(R)p(P)p(U')$$

are operations of type VIII associated respectively with the Jacobian planes $xp'u', x'p'u, x'pu'$, leaving unaltered, respectively, the nodes V', V, V'. Thus the operation T arises equally from any one of the Jacobian planes. The operation T is easily seen to permute the nodes in Σ according to the permutation

$$(XX'YZ'ZY')(PR'QP'RQ')(UW)(U'W'),$$

so that it leaves V and V' unaltered, interchanges the remaining nodes on the edges u, u', and interchanges the other two pairs of opposite edges of the tetrahedron, permuting the six nodes on the edges of each pair in a cycle of 6. There are two distinct operations of this type (inverse to each other) keeping fixed any one of the 27 pairs of nodes such as V, V' lying on opposite edges of the tetrahedron. Thus the 54 operations so arising are distinct. Further, T interchanges the Jacobian planes ([illegible]$pu, x'p'u$) and $(xp'u', x'pu')$ and permutes the three cross-planes of Σ which contain V and V', namely, the planes containing

$$VV'XZ'QQ', \quad VV'ZX'PP', \quad VV'YY'RR',$$

in cyclic order. Whence it is easily seen that two distinct Steiner primes cannot give rise to the same operation.

By elementary transformations it is found that

$$\begin{aligned} T^2 &= p(X)p(Y)p(U)p(U')\cdot p(X)p(Y)p(U)p(U') \\ &= p(X)p(Y)p(U)p(U')\cdot p(U)p(U')p(X')p(Z') = p(X)p(Y)p(X')p(Z'), \end{aligned}$$

an operation of period 3 and type XI, and that.

$$\begin{aligned} T^3 &= p(X)p(Y)p(X')p(Z')\cdot p(X)p(Y)p(U)p(U') \\ &= [p(X)p(Y)]^2 p(X')p(Z')p(U)p(U') \\ &= p(Y)p(X)p(X')p(Z')p(U)p(U') = p(Y)p(X)p(U)p(U')p(X)p(Y) \\ &= p(V)p(V'), \end{aligned}$$

an operation of period 2 and type III. Thus T is of period 6.

It is easy to see that the subgroup of G which keeps Σ fixed contains operations which transform any pair of opposite edges of the tetrahedron into any other pair, and operations (products of projections from suitably chosen nodes on u and u') which keep u and u' unchanged and transform V, V' into any other assigned pair of nodes on these edges. Thus every one of the 54 operations T derived from Σ is conjugate to one of the two operations T, T^{-1} which keep V and V' fixed. And T and T^{-1} are conjugate, for

$$T = p(X)p(Y)p(U)p(U'), \quad T^{-1} = p(U')p(U)p(Y)p(X) = p(Y)p(X)p(W')p(W),$$

and

$$\begin{aligned} [p(Y)p(Y')]\,T[p(Y)p(Y')]^{-1} &= p(Y)p(Y')p(X)p(Y)p(U)p(U')p(Y')p(Y) \\ &= p(Y)p(X)p(Y)p(Q')p(Q)p(Y) \\ &= p(Y)p(X)p(W')p(W) = T^{-1}. \end{aligned}$$

Thus all the 54 operations derived from Σ are conjugate. And since G is transitive on the Steiner primes the operations of this type corresponding to different primes are all conjugate in G. Thus G contains

XII. *A conjugate set of* 2160 *operations of period* 6, *products of four projections, whose squares are of type XI and whose cubes are of type III.*

The next type of product that arises can be put in the form $T_1 \cdot p(A)$ where, again, T_1 is an operation of type VIII associated with one of the Jacobian planes in Σ leaving a node B in this plane invariant, and A is a node in Σ, not lying in the Jacobian plane, such that A and B lie on intersecting edges of the tetrahedron. For each operation T_1 there are thus six possible positions for A. Hence there are, in all,

$4.18.6 = 432$ products of this nature to be considered. I shall prove that these lead to 144 distinct operations.

A typical operation of this type is $T = p(X)p(Y)p(U)p(X')$. By elementary transformations it is seen that

$$\begin{aligned} p(X)p(Y)p(U)p(X') &= p(X)p(Y)p(R')p(U) = p(R')p(U')p(V')p(U) \\ &= p(U')p(V')p(Q')p(U) \\ &= p(U')p(V')p(Y')p(Q') = p(Y')p(Q)p(R)p(Q') \\ &= p(Q)p(R)p(X')p(Q'). \end{aligned}$$

Thus T is expressible in the form $T_1 . p(A)$ in three ways, in which the operation T_1 belongs to the respective Jacobian planes $xpu, xp'u', x'pu'$. Hence the 432 products give rise to at most 144 distinct operations. The operation T permutes the 18 nodes lying in Σ in the cycles

$$(XQV'ZPU'YRW')(X'Q'U)(Y'R'W)(Z'P'V).$$

From this it is easily seen (i) that no cross-plane lying in Σ is unaltered by T, (ii) that the three Jacobian planes in Σ other than $x'p'u$ are permuted cyclically, (iii) that the Jacobian plane $x'p'u$ is left unchanged, but that the three triads of κ-lines

$$(X'Q'W, Y'R'V, Z'P'U), \quad (Q'UY', R'WZ', P'VX'), \quad (UX'R', WY'P', VZ'Q'),$$

which are the edges, lying in this plane, of the tetrahedra of Jacobian planes in the three Steiner primes other than Σ which contain the plane $x'p'u$, are permuted cyclically. From these properties it is easy to see that distinct Steiner primes give distinct operations.

It will now be proved that there are just 144 distinct operations of this type derived from Σ. In the first place it is to be observed that

$$\begin{aligned} T^3 &= p(X)p(Y)p(U)p(X').p(X)p(Y)p(U)p(X').p(X)p(Y)p(U)p(X') \\ &= [p(X)p(Y)]^2.p(V)p(X')p(U)p(X').p(X)p(Y)p(U)p(X') \\ &= p(Y)p(X).p(X)p(Y)p(W)p(P')p(U)p(X') \\ &= p(W)p(P')p(U)p(X'), \end{aligned}$$

which is easily recognized as an operation of type VII and period 3 associated with the Jacobian plane $x'p'u$. Thus, since T contains a cycle of period 9, T is an operation of period 9. The three operations T, T^4, T^7 effect the same permutation of the nine nodes in the plane $x'p'u$, this permutation consisting of three cycles each of three non-collinear nodes. It is easy to see that the three operations $p(X)p(Y)p(U)p(A)$, where $A = X'$, Y' or Z', the transforms of these operations by $p(X')$, and the squares of these six operations (which are all of the same type† as T), effect 12 different permutations of the nine nodes in $x'p'u$. Thus these 12 operations, together with their fourth and seventh powers (which may be verified to be of the same type‡) give 36 operations of this type, all distinct, which leave the plane $x'p'u$ unchanged. Thus, from the four planes $xpu, xp'u', x'pu', x'p'u$ we get 144 distinct operations.

---

† In fact, $T^2 = p(Y)p(X)p(V)p(R')$.

‡ In fact, $T^4 = p(X)p(Y)p(V)p(Z')$, $T^7 = p(X)p(Y)p(W)p(Y')$.

These operations cannot all be conjugate in G, for their cubes, which are of type VII, belong to two distinct conjugate sets. It will be proved, in fact, that they form two conjugate sets in G. In the first place, since

$$p(X)p(Y)p(U) = p(U)p(R)p(P) = p(R)p(P)p(W),$$

the operations $p(X)p(Y)p(U)p(X')$ and $p(R)p(P)p(W)p(X')$ are identical. It follows that every operation $T_1.p(A)$ of the type considered can be expressed as a product $p(X_1)p(Y_1)p(U_1)p(X_1')$ of projections from four nodes X_1, Y_1, U_1, X_1' lying in a Steiner prime Σ_1 with the property that X_1Y_1 is an edge of the tetrahedron formed by the Jacobian planes in Σ_1, U_1 lies in one of these planes through X_1Y_1, and X_1' lies on the polar κ-line of X_1Y_1. Now there is a unique automorphism of the set of nodes in which X, Y, U, X', correspond respectively to X_1, Y_1, U_1, X_1'; for incidence relations determine uniquely (i) the node corresponding to the third node Z on the κ-line XY, (ii) the nodes corresponding to the remaining nodes in the Jacobian plane XYU, (iii) the nodes corresponding to the remaining nodes in Σ, (iv) the nodes corresponding to the remaining 27 nodes, which are the poles of the Jordan primes containing the various cross-planes in Σ. Hence any two products of the type $T_1.p(A)$ are conjugate in the group G^* of automorphisms of the set of nodes. Since G is a subgroup of index 2 in G^* these operations form at most two conjugate sets in G. But it has been seen that they cannot all belong to the same conjugate set. Thus they form exactly two conjugate sets, and hence G contains

XIII. 5760 *operations of period* 9, *products of four projections, forming two conjugate sets of* 2880 *operations each, whose cubes are of type VII.*

It is not difficult to prove that any irreducible product of four projections from nodes of Σ which do not lie in a plane can be reduced by elementary transformations to one of the three types XI, XII, XIII. Any other irreducible product thus contains at least five factors. There is, in fact, just one further type of product which is irreducible, namely a product of five projections expressible in the form $T_1.p(A)$, where T_1 is an operation of type VII, the product of projections from four nodes in a Jacobian plane lying in Σ, which leaves every node in this plane invariant, and A is a node in Σ which does not lie in this plane. Since each Jacobian plane determines two such operations T_1 and there are nine nodes of Σ not lying in any given plane, there are 72 products of this type to be considered. In fact these give rise to 36 distinct operations.

Consider the typical product $T = p(X)p(Y)p(P)p(Q)p(X')$. It is easily seen that this operation leaves invariant the four nodes X, Y, Z, Z', interchanges X' and Y', and permutes the remaining nodes in Σ in two cycles of six, namely,

$$(PW'QU'RV')(P'VR'UQ'W).$$

Thus T leaves invariant the two Jacobian planes $x'pu', x'p'u$ which pass through x' and interchanges the two Jacobian planes $xpu, xp'u'$ which pass through the line x

containing three invariant nodes. Considerations of a nature similar to those previously employed show that distinct Steiner primes must give rise to distinct operations.

Since
$$T = p(X)p(Y)p(P)p(Q)p(X') = p(X)p(Y)p(X')p(V')p(W')$$
$$= p(X)p(Y)p(V')p(W')p(Y'),$$

and $p(X)p(Y)p(V')p(W')$ is an operation of type VII associated with the Jacobian plane $xp'u'$, each operation of this type is expressible in at least two ways in the form $T_1.p(A)$. On the other hand, the two operations T and T^{-1} (which are distinct since they effect distinct permutations of the twelve nodes on the edges p, p', u, u' of the tetrahedron) keep fixed X, Y, Z, Z', that is, they keep fixed Z' and each node of the polar κ-line of the edge of the tetrahedron on which Z' lies, and by symmetry a similar pair of operations is associated with each of the 18 nodes in Σ. There are thus exactly 36 distinct operations of this type derived from each Steiner prime.

Since
$$p(X)p(Y)p(P)p(Q) = p(R)p(P)p(X)p(Y),$$

it follows that

$$\begin{aligned} T^2 &= p(R)p(P)p(X)p(Y)p(X').p(X)p(Y)p(P)p(Q)p(X') \\ &= p(R)p(P)[p(X)p(Y)]^2.p(X')p(P)p(Q)p(X') \\ &= p(R)p(P)p(Y)p(X)p(V')p(W') \\ &= p(R)p(P)p(Y)p(Q')p(P')p(X) \\ &= p(R)p(P)p(Q')p(P')p(X)p(X) = p(R)p(P)p(Q')p(P'), \end{aligned}$$

an operation of type XI and period 3, and

$$\begin{aligned} T^3 &= p(X)p(Y)p(P)p(Q)p(X').p(R)p(P)p(Q')p(P') \\ &= p(X)p(Y)p(P)p(Q)p(R)p(P).p(Z')p(Q')p(P') \\ &= p(X)p(Y)p(Q)p(P).p(V)p(U)p(Z') \\ &= p(X)p(Q)p(W)p(P)p(W)p(V)p(Z') \\ &= p(X)p(Q)p(X)p(V)p(Z') \\ &= p(Z'), \end{aligned}$$

so that T is of period 6 (and incidentally does not belong to a type already considered).

It has been seen that T can be expressed in the form $T_1.p(A)$, where T_1 is an operation of type VII in just two ways, the two operations T_1 being

$$p(X)p(Y)p(P)p(Q) \quad \text{and} \quad p(X)p(Y)p(V')p(W'),$$

and the corresponding projections $p(A)$ being $p(X')$ and $p(Y')$. Since

$$p(Z').p(X').p(Z') = p(Y')$$

and

$$p(Z').p(X)p(Y)p(P)p(Q).p(Z') = p(X)p(Y)p(U')p(V') = p(X)p(Y)p(V')p(W'),$$

it is seen that the two operations T_1 which give rise to T, and the two corresponding projections $p(A)$, are transformed into each other by the same operation of G. Hence if $T_1'.p(A')$ is any other operation of this type which is conjugate to T, then T_1' must be conjugate to T_1. Since the operations of type VII form two conjugate sets in G there must therefore be at least two conjugate sets of operations of the type under consideration. But, by reasoning similar to that used in discussing the operations of type XIII, it is easy to see that any two operations of this type are conjugate in G^*, the group of automorphisms of the set of nodes. Hence the number of conjugate sets is seen to be exactly two, and hence G contains

XIV. 1440 *operations of period* 6, *products of five projections, forming two conjugate sets of* 720 *operations each, whose squares are of type VII and whose cubes are projections.*

Every product of projections from a set of nodes lying in Σ belongs to a type already considered. For the projection from any node in Σ leaves Σ unchanged, so that the group generated by the projections from the 18 nodes in Σ either coincides with, or is a subgroup of, the group of order 648 consisting of all the operations of G leaving Σ invariant. In fact, it coincides with this group of order 648, and its operations consist of (i) identity, (ii) 18 projections from nodes in Σ, (iii) 27 operations of type III arising from e-lines in Σ, (iv) 84 operations of type IV arising from κ-lines, (v) 36 operations of type V arising from d-planes, (vi) 162 operations of type VI arising from c-planes, (vii) 8 operations of type VII and 72 of type VIII arising from j-planes, and (viii) 6 operations of type XI, 54 of type XII, 144 of type XIII, and 36 of type XIV; making 648 operations in all. Thus every product of projections from a set of nodes in Σ has been accounted for.

6. THE REMAINING OPERATIONS OF G, AND THE SUBGROUP LEAVING A NODE INVARIANT

Now return to the discussion of the subgroup of G, which will be denoted by G_5 (in accordance with the notation of § 7 below), leaving invariant a node O and the Jordan prime Π which is the polar of O. This group is of order 576. It was proved, in § 5·2, that the projections from the nodes in Π generated a subgroup of G_5, of order 192, and the operations of this subgroup were enumerated in nine sets (i) to (ix). The other operations of known types which belong to G_5 are easily seen to be: (x) 16 operations of type VII, two associated with each Jacobian plane through O; (xi) 16 operations of type VIII, two associated with each Jacobian plane through O, which are the products of $p(O)$ with the operations of the previous set; (xii) 32 operations of type XI; (xiii) 96 operations of type XII; (xiv) 128 operations of type XIV, falling into two sets, of 32 and 96 operations respectively, which are the products of $p(O)$ with the operations of the two preceding sets. In each of the last three cases, the number of operations in the set is $nm/45$, where n is the total number of operations of the type considered and m is the number of nodes left unchanged by an operation of this type. The division of the 128 operations of the set (xiv) into sets of 32 and 96 arises from the fact that of the four nodes left invariant by an operation of type XIV a particular set of three lies on a κ-line.

These fourteen sets account for 480 operations of G_5. It will be proved that the remaining 96 operations belong to a new type, and that this type, together with those obtained previously, exhausts the operations of G.

Since $p(O)$ leaves each node of Π invariant (and is the only operation of G other than identity with this property) the operations of G_5 permute the 12 nodes of Π according to a group of order 288. The 480 operations of G_5 listed in the sets (i) to (xiv) account for 240 of these permutations, and it will be convenient to list typical permutations of each type that arises. The notation of § 5·2 is used for the nodes in Π, and the types of permutation are referred to as (*a*), (*b*), ..., (*i*); with an indication, for each type of permutation, of those of the sets (i), (ii), ..., of operations of G_5 from which the permutation arises. The details of the calculation (which are elementary) are left to the reader. The 240 permutations of the nodes in Π are

(*a*) the identical permutation [(i), (ii)];

(*b*) 12 permutations such as $(A_2A_3)(B_2B_3)(C_2C_3)(D_2D_3)$, which keep fixed all the vertices of one tetrahedron and interchange in pairs the vertices of the other tetrahedra collinear with one vertex of the former (in this case the vertex A_1) [(iii),(iv)];

(*c*) 9 permutations such as $(A_2B_2)(C_2D_2)(A_3B_3)(C_3D_3)$ which leave invariant the vertices of one tetrahedron and interchange the vertices of a pair of opposite edges of each of the other two, the two edges of one of these tetrahedra meeting the two edges of the other [(v)];

(*d*) 32 permutations such as $(A_1A_2A_3)(B_1B_2B_3)(C_1C_2C_3)(D_1D_2D_3)$ permuting the tetrahedra cyclically, one cycle of three nodes (here $A_1A_2A_3$) lying on a κ-line and the others in the three cross-planes passing through the κ-line [(vi), (vii)];

(*e*) 36 permutations such as $(A_1B_3B_1A_3)(C_1D_3D_1C_3)(A_2B_2)(C_2D_2)$ which interchange in pairs the vertices on a pair of opposite edges of one of the tetrahedra and permute cyclically the other nodes in Π which lie in two cross-planes, one through each of the edges, in such a way that $A_1B_3, B_3B_1 \dots$ are κ-lines [(viii)];

(*f*) 6 permutations such as $(A_1C_1)(B_1D_1)(A_2B_2)(C_2D_2)(A_3D_3)(B_3C_3)$ which interchange the nodes on three mutually skew pairs of opposite edges of the tetrahedra [(ix)];

(*g*) 16 permutations such as $(A_1B_2B_3)(B_1D_2C_3)(C_1C_2A_3)(D_1A_2D_3)$ permuting the tetrahedra cyclically, the four cycles of three nodes belonging to four mutually skew κ-lines, whose polars lie in a Jacobian plane through O [(x), (xi)];

(*h*) 32 permutations such as $(B_1D_1C_1)(A_2C_2D_2)(A_3C_3D_3)$ keeping fixed the nodes on a κ-line (here $A_1B_2B_3$) and permuting cyclically the three other vertices of each tetrahedron [(xii), (xiv)];

(*i*) 96 permutations such as $(A_1B_2)(B_1D_2D_1A_2C_1C_2)(A_3D_3C_3)$ which keep fixed a vertex (here B_3) of one tetrahedron, permute cyclically the other vertices of this tetrahedron, interchange two vertices of the other tetrahedra lying on a κ-line through the fixed vertex, and permute the other nodes in a cycle of six in such a way that consecutive edges of the hexagon $B_1D_2D_1A_2C_1C_2$ pass through

$$C_3, A_3, D_3, C_3, A_3, D_3$$

where the cyclic order is the same as that in which A_3, D_3, C_3 are permuted [(xiii), (xiv)].

It will be proved that the remaining operations of G_5 consist of:

(xv) 96 operations of the form T_1T_2, where T_1 is a product of projections from three non-collinear nodes in a Jacobian plane through O which leaves O invariant, and T_2 is a product of projections from two vertices of one of the three tetrahedra is Π, the product of any such operation with $p(O)$ being of the same type.

Further, it will be shown that the permutations effected by these operations on the nodes in Π consist of:

(j) 48 permutations of the form $(A_1B_2D_3B_1D_2A_3)(C_1C_2C_3D_1A_2B_3)$ permuting the vertices of the three tetrahedra in two cycles of six, the square of the permutation being a permutation of the set (g).

In the first place, if T_2 is the product of two projections from nodes on an edge of one of the tetrahedra, then $T_2p(O)$ is the product of projections from the two nodes of the opposite edge. Thus if T_1T_2 is an operation of the type considered, so also is $T_1T_2p(O)$. Moreover, the permutations effected by T_1 and T_2 on the nodes in Π have the forms (g) and (c) respectively, and the product of the typical permutations given for these forms is the permutation given for the form (j). An examination of the nature of the permutations (g) and (c) readily shows that the two permutations chosen are typical of any such pair. Moreover, it is easily seen that precisely 48 permutations of type (j) exist, three having any given permutation of the form (g) as square; and this indicates the existence of 96 operations of the set (xv), two giving rise to each of the 48 permutations of the nodes in Π, these being of the forms $T_1T_2, T_1T_2p(O)$.

Further study of the operations of this type, which now follows, makes it clear that the operations are completely symmetrical with respect to the three tetrahedra, and that each product T_1T_2 can be expressed in three equivalent forms, in which T_2 is associated with each of the three tetrahedra in turn. A direct verification of this, by means of elementary transformations, is left to the reader. The relation of these operations to the whole set of nodes will now be considered in detail. Take, as a typical operation of this type, the product

$$L = p(14)p(12, 56, 34)p(16, 35, 24).p(16)p(16, 45, 23),$$

where the first three factors form an operation T_1 of type VIII keeping the node (45) and a Jacobian plane through it fixed, and the last two factors form an operation T_2, the product of projections from two nodes in the polar prime of (45). By means of elementary transformations the product can be transformed into a rather more convenient form. Thus

$$\begin{aligned} &p(14)p(12,56,34)p(16,35,24)p(16)p(16,45,23)\\ &\qquad = p(46)p(41)p(15,34,26)p(16,35,24)p(16,45,23)\\ &\qquad = p(46)p(41)p(15,34,26)p(34)p(16,35,24)\\ &\qquad = p(46)p(41)p(43)p(15,34,26)p(16,35,24), \end{aligned}$$

and so $$L = p(46)p(41)p(43)p(15,34,26)p(16,35,24),$$

which will be regarded henceforth as the standard form. Now

$$\begin{aligned} L^2 &= p(46)p(41)p(43)p(15,34,26)p(16,35,24)\\ &\qquad \times p(46)p(41)p(43)p(15,34,26)p(16,35,24)\\ &= [p(46)p(41)p(43)]^2 . p(12,35,46)p(13,45,26)p(15,34,26)p(16,35,24)\\ &= p(16)p(36) . p(41)p(43) . p(12,35,46)p(16,35,24)p(14,56,23)p(14,36,25)\\ &= p(36)p(31)p(31)p(41) . p(26)p(12,35,46)p(14,56,23)p(14,36,25)\\ &= p(36)p(41)p(26)p(14,56,23)p(12,35,46)p(14,36,25)\\ &= p(26) . p(23)p(41)p(14,56,23) . p(12,35,46)p(14,36,25)\\ &= p(26)p(56) . p(14,23,56)p(12,35,46)p(14,36,25)\dagger\\ &= p(25)p(26)p(14,23,56)p(14,36,25)p(12,35,46)\\ &= p(25)p(14,23,56)p(12,35,46), \end{aligned}$$

which is easily recognized as an operation of type VIII leaving (45) invariant, and

$$\begin{aligned} L^3 &= p(46)p(41)p(43)p(15,34,26)p(16,35,24) . p(25)p(14,23,56)p(12,35,46)\\ &= p(46)p(41)p(43)p(25) . p(12,34,56)p(16,23,45)p(14,23,56) . p(12,35,46)\\ &= p(16)p(36)p(25)p(46)p(16,23,45)p(14,23,56)p(26)p(12,35,46)\\ &= p(16)p(36)p(25)p(16,23,45)p(26)p(12,35,46)\\ &= p(16)p(36)p(25)p(26)p(12,36,45)p(12,35,46)\\ &= p(16)p(36)p(25)p(26)p(56)p(12,36,45)\\ &= p(16)p(36)p(26)p(12,36,45), \end{aligned}$$

which is an operation of type X whose square is the projection $p(45)$. Thus L is an operation of period 12, and $L^6 = p(45)$.

The operation L keeps the node (45) fixed; it permutes the nodes in the polar prime of (45) in two cycles of six, and the remaining 32 nodes in two cycles of four and two cycles of twelve, the actual cycles being

[(12), (16), (13, 26, 45), (36), (16, 45, 23), (13)],
[(23), (26), (12, 45, 36), (16, 23, 45), (13, 45, 26), (12, 36, 45)],

[(24), (12, 34, 56), (25), (12, 35, 46)], [(14, 23, 56), (15, 34, 26), (15, 23, 46), (14, 35, 26)],

[(14), (15, 26, 34), (13, 56, 24), (12, 56, 34), (16, 25, 34), (14, 56, 23), (15),
(14, 26, 35), (13, 46, 25), (12, 46, 35), (16, 24, 35), (15, 46, 23)],

[(34), (16, 34, 25), (15, 24, 36), (15, 36, 24), (13, 25, 46), (46), (35), (16, 35, 24),
(14, 25, 36), (14, 36, 25), (13, 24, 56), (56)].

---

† Since the nodes (14), (23), (56), (14, 23, 56), (14, 56, 23) form a Jordan pentahedron.

Since L keeps only one node fixed the operations of this type arising from different nodes are distinct. Since the squares of these operations are of type VIII, the aggregate of operations of this type forms at least two conjugate sets in G. It will be proved that there are exactly two such sets. For the eight nodes belonging to the two cycles of four lie in a Jacobian plane through (45). The four nodes in any one of these cycles lie alternately on two κ-lines through (45), and the joins of consecutive nodes pass, in order, through the four nodes of the other cycle. Thus either cycle determines the other. There are six pairs of cycles of this kind associated with each Jacobian plane through (45), and it is easy to see that there is a collineation in G which transforms any one of them into any other. Thus, of the 96 operations such as L which keep (45) fixed, there are two associated with each of the 48 pairs of cycles, and if one of these is L the other is clearly L^5. Then every operation of the same type as L is conjugate to one of the two operations L, L^5. But these are not conjugate, since their squares are non-conjugate operations of type VIII. Thus G contains

XV. 4320 *operations of period* 12, *products of five projections, forming two conjugate sets of* 2160 *operations each, whose squares are of type* VIII *and whose cubes are of type* X.

Since the types I to XV include 25920 distinct operations, all the operations of G are accounted for, and it follows that *every operation of G is expressible as the product of five or fewer projections*. The distribution of the operations of G in conjugate sets agrees with that obtained, by very different methods, by Frame (1936, p. 483).

7. Some subgroups of G

As an illustration of the usefulness of the classification of operations of G undertaken above it will be shown how the nature of the operations of five important maximal subgroups of G can be determined. These five subgroups are:

(i) the subgroup G_1 of order 960 leaving invariant a Jordan pentahedron,
(ii) the subgroup G_2 of order 720 leaving invariant the nodes of a 15-set,
(iii) the subgroup G_3 of order 648 leaving invariant a Steiner prime,
(iv) the subgroup G_4 of order 648 leaving invariant a Jacobian plane,
(v) the subgroup G_5 of order 576 leaving invariant a node.

Of these, G_3 and G_5 have been described in §§ 5·3 and 6 respectively. The procedure for determining the operations of the other subgroups will be given in outline, leaving the details to the reader and referring him to the table at the end of the section for the results.

For G_1 one observes first that there are 192 operations which transform a Jordan pentahedron into itself and keep a node of this fixed, these forming a subgroup of index 3 in the G_5 which keeps the node fixed. These operations are easily identified in the list of operations of G_5 already given. By examining the permutations effected on the nodes of the Jordan pentahedron by these operations it is easy to see, first, that from the five vertices of the pentahedron 576 distinct operations are obtained in this way, and secondly, that these operations all effect *even* permutations of the five nodes of the pentahedron. Hence, as $576 > \frac{1}{2}.960$, every operation of G_1 permutes the vertices of the pentahedron evenly. Any of the remaining 384 operations of G_1 thus effects an even permutation of the nodes of the pentahedron which keeps no vertex fixed. Such a permutation must therefore consist of a single cycle of five vertices, and the corresponding operation of G_1 must be of type IX, the only type whose period is divisible by 5.

For G_2 it is sufficient to observe that, from a consideration of the 15-set consisting of the nodes (ij), the group is simply isomorphic with the symmetric group of degree 6, and is generated by the projections $p(ij)$ which correspond to transpositions. The only operation of G_2 whose nature is not at once clear is the cyclic operation of period 6, and it may be seen, either by reducing the product of projections corresponding to such an operation, or (more simply) by considering the natures of the square and cube of the operation, that it is actually of type XII.

For G_4 there have been given in § 4·3 54 operations generated by projections from the nodes lying in a Jacobian plane, and from the list of operations of G_3 given in § 5·3 one can easily select a further 108 operations, associated with each Steiner prime through the plane, which leave the plane invariant. These account in all for 486 operations of G_4, and the remaining operations are easily seen to consist of 108 operations of type XV and 54 of type X which are the cubes of these.

In table 1 are enumerated the operations of each of the fifteen types in G and its subgroups G_1, G_2, G_3, G_4, G_5, and also, for each type of operation, its period p and the least number n of projections whose product is of the given type. Each type of operation corresponds to a single conjugate set in G with the exception of those marked by an asterisk, which each yield two conjugate sets.

TABLE 1. LIST OF OPERATIONS OF THE SIMPLE GROUP G AND ITS SUBGROUPS G_1, G_2, G_3, G_4, G_5

			number of operations of type in					
type	p	n	G	G_1	G_2	G_3	G_4	G_5
I	1	—	1	1	1	1	1	1
II	2	1	45	5	15	18	9	13
III	2	2	270	70	60	27	—	30
IV	3	2	480	—	40	84	48	32
V	6	3	1440	—	120	36	72	32
VI	4	3	3240	120	180	162	—	72
VII*	3	4	80	—	—	8	26	16
VIII*	6	3	720	—	—	72	90	16
IX	5	4	5184	384	144	—	—	—
X	4	4	540	60	—	—	54	12
XI	3	4	240	80	40	6	24	32
XII	6	4	2160	80	120	54	—	96
XIII*	9	4	5760	—	—	144	144	—
XIV*	6	5	1440	160	—	36	72	128
XV*	12	5	4320	—	—	—	108	96
			25920	960	720	648	648	576

8. SOME SETS OF GENERATORS OF G

I shall conclude by showing how the expressions of the operations of G in terms of projections can be usefully employed in investigating various systems of generators of the group. It does not seem unreasonable to suppose that a systematic investigation of such sets could be made along these lines; but no such investigation (which would of necessity be elaborate) has yet been made, and I content myself with a few illustrative examples, some of them new and others confirming results already obtained.

The group G is generated by the 45 projections (or, indeed, as it is simple, by the operations of any one of its conjugate sets). But these are connected by many relations, and it is natural to inquire as to the minimum number of projections in terms of which all the others can be expressed (and hence any operation of G can be expressed). This number must be five at least, since any four nodes determine a prime, and it has already been seen that the projections from the nodes which lie in any prime generate a proper subgroup of G. In fact, five projections suffice. This is shown by proving that any projection can be expressed in terms of the projections from the five nodes

$$(12),\quad (16),\quad (25),\quad (46),\quad (15,26,34),$$

(which are clearly typical of numerous other similar sets). The proof rests on the identities

$$p(ij)\,p(jk)\,p(ij) = p(ik),$$
$$p(ij)\,p(ik,jl,mn)\,p(ij) = p(jk,il,mn),$$
$$p(ij)\,p(kl)\,p(mn) = p(ij,kl,mn)\,p(ij,mn,kl),$$

which are consequences of the fundamental relations (1) between the projections given in § 2 and the relations of incidence between the nodes. From the first of these it is easy to see that every $p(ij)$ for which $i,j \neq 3$ is expressible in terms of $p(12)$, $p(16), p(25), p(46)$—in fact these nodes (ij) are just the nodes of an n-prime. From the second identity it easily follows that every $p(ij,kl,mn)$ is expressible in terms of these ten projections and $p(15,26,34)$. Whence, from the third identity, $p(i3)$ can be found, proving the theorem.

Coxeter has shown (Coxeter 1940, p. 474) that the relations

$$U^3 = V^2 = W^2 = X^2 = Y^2 = (UV)^3 = (VW)^3 = (WX)^3 = (VY)^3$$
$$= (UW)^2 = (UX)^2 = (UY)^2 = (VX)^2 = (WY)^2 = (XY)^2 = 1$$

constitute an abstract definition for G in terms of five suitably chosen operations U, V, W, X, Y. These generating relations are satisfied by taking

$$U = p(23)p(12)p(56)p(45),\quad V = p(25)p(36),\quad W = p(13)p(45),$$
$$X = p(23)p(56),\quad Y = p(16,25,34)p(15,36,24).$$

Here we recognize U as an operation of type XI and period 3, and V, W, X, Y as operations of type III and period 2. That the other generating relations are satisfied is most simply shown by reducing the products of pairs of these operations to one of the standard types by means of elementary transformations. Two examples of the process will suffice. Thus

$$\begin{aligned} XY &= p(23)\,p(56)\,p(16,25,34)\,p(15,36,24) = p(15,36,24)\,p(23)\,p(56)\,p(15,36,24) \\ &= p(15,26,34)\,p(16,35,24), \end{aligned}$$

an operation of type III and period 2, and

$$\begin{aligned} UV &= p(23)\,p(12)\,p(56)\,p(45)\,p(25)\,p(36) = p(13)\,p(23)\,p(56)\,p(25)\,p(24)\,p(36) \\ &= p(13)\,p(23)\,p(25)\,p(26)\,p(36)\,p(24) = p(13)\,p(35)\,p(26)\,p(24), \end{aligned}$$

an operation of type XI and period 3. The remaining relations may be verified in a similar way.

That these five operations generate G follows from the stronger theorem that G is generated by the two operations $H = UVW$ and $K = XY$. That G possesses a *pair* of generating operations seems to have been first pointed out by Brahana (1930, p. 533), and the identification of Brahana's operations with H and K is stated, without proof, by Coxeter (1940, p. 475). A simple proof of the fact that H and K generate G will now be given. Since

$$\begin{aligned} UVW &= p(13)\,p(35)\,p(26)\,p(24)\,.\,p(13)\,p(45) = p(15)\,p(26)\,p(24)\,p(45) \\ &= p(42)\,p(46)\,p(15)\,p(45) = p(42)\,p(46)\,p(45)\,p(41), \end{aligned}$$

an operation of period 5 and type IX, it follows that

$$H = p(42)\,p(46)\,p(45)\,p(41), \quad K = p(15,26,34)\,p(16,35,24),$$

and

$$\begin{aligned} HK &= p(42)\,p(46)\,p(45)\,p(41)\,.\,p(15,26,34)\,p(16,35,24) \\ &= p(46)\,p(41)\,p(26)\,p(15)\,p(15,26,34)\,p(16,35,24) \\ &= p(46)\,p(41)\,p(43)\,p(15,34,26)\,p(16,35,24),\dagger \end{aligned}$$

which is the operation L of type XV and period 12 considered in § 6. Thus the group generated by H and K contains L, and hence contains $L^6 = p(45)$. It was shown in § 2 that if an operation T of G transforms a node A into a node A', then

$$Tp(A)\,T^{-1} = p(A').$$

Hence, since H permutes the nodes in such a way that the digits in their symbols undergo the permutation (15624), the operations $H^r L^6 H^{-r}$, for $r = 1, 2, 3, 4$ are respectively $p(16), p(25), p(46), p(12)$. And hence $p(14) = p(16)\,p(46)\,p(16)$ belongs to the group generated by H, K, and so does $Lp(14)\,L^{-1}$, which is $p(15, 26, 34)$, since L transforms the node (14) into (15, 26, 34) (from the expression of L as a permutation of the nodes given in § 8). Thus the group generated by H and K contains the projections from the five nodes (12), (16), (25), (46), (15, 26, 34), and it was seen at the beginning of the present section that these projections generate the entire group.

---

† Since (15), (26), (34), (15, 26, 34), (15, 34, 26) form a Jordan pentahedron.

It is possible to show that G can also be generated by the two operations L and $p(A)$, where A is one of the nodes belonging to a cycle of twelve permuted by L. It is easy to verify that, in the first of the two cycles listed in § 6, the join of any two consecutive nodes in the cycle is a κ-line, whose third node is one of those permuted by L in a cycle of six, and that in the operation L^7 the second cycle has this property. Since $(L^7)^7 = L^{49} = L$ it follows that there is nothing lost in generality by supposing that A belongs to the first cycle. And since the nodes in any one cycle are permuted cyclically by L, the projections from these nodes are all of the form L^rTL^{-r}, where T is any one of them. It is thus sufficient to prove that L and $p(14)$ generate the group. From what has been said, the group generated by L and $p(14)$ contains the projections from all 12 nodes of the cycle to which (14) belongs. And from the projections from two consecutive nodes in the cycle can be deduced the projection from the third node on the κ-line joining them, which is a node belonging to the second of the two cycles of six. And the projection $L^6 = p(45)$ also belongs to the group. Thus, in particular, the group generated by L and $p(14)$ contains the projections from the nodes (23), (26) and hence the projections from (36) and the remaining nodes of the first cycle of six to which (36) belongs. Thus the group contains the five projections $p(12), p(14), p(23), p(26), p(15, 26, 34)$, and these are easily seen to form a set of generators of G similar to that given at the beginning of the section.

It will be noticed that

$$\begin{aligned} Lp(14) &= p(46)p(41)p(43)p(15,34,26)p(16,35,24)p(14) \\ &= p(46)p(13)p(13,26,45)p(12,46,35), \end{aligned}$$

and it is easy to see that the four nodes concerned in these four projections determine an n-prime, so that $Lp(14)$ is an operation of period 5 and type IX. The operation $L^{-1}p(14)$, whose inverse $p(14)L$ is conjugate to $Lp(14)$, is of the same type. Also,

$$\begin{aligned} L^7p(14) = L^6.Lp(14) &= p(45).p(46)p(13)p(13,26,45)p(12,46,35) \\ &= p(56)p(45)p(13)p(13,26,45)p(12,46,35) \\ &= p(56)p(26)p(13,45,26)p(12,46,35)\dagger \\ &= p(56)p(26)p(12,46,35)p(13,45,26). \end{aligned}$$

The product of the first three factors is easily seen to be an operation of type VIII, associated with a Jacobian plane containing the nodes (56), (26), (12, 46, 35), which keeps fixed the node (16, 45, 23). And since this node is joined by a κ-line to (13, 45, 26) it follows that $L^7p(14)$ is an operation of type XIII and period 9. So, also, $L^5p(14)$, which is conjugate to $p(14)L^5 = [L^7p(14)]^{-1}$, is of period 9.

Hence $Lp(14)$ and $L^7p(14)$ have periods 5 and 9. It is easily seen that the same is true if $p(14)$ be replaced by the projection from any other node in the cycle of twelve

---

† Since (13), (26), (45), (13, 26, 45), (13, 45, 26) form a Jordan pentahedron.

to which (14) belongs, and that for projections from nodes of the other cycle of twelve the corresponding periods are 9 and 5.

Since L and $p(14)$ generate G so also do either of the pairs of operations $Lp(14), p(14)$ or $L^7p(14), p(14)$, that is, G admits a pair of generators of periods 5, 2 or 9, 2 whose product is of period 12. Frame has given, as a pair of generators of G, two operations of periods 9 and 2 whose product has period 5 (Frame 1936, p. 482). It is easy to find such a pair. In fact, G is generated by the two operations $L^7p(14)$ and $L^6 = p(45)$, whose product $Lp(14)$ has period 5. To see this, write

$$P = L^7p(14), \quad Q = L^6, \quad R = QP = Lp(14).$$

Then R belongs to the group generated by P and Q. Moreover,

$$P^{-1}R = p(14)\,L^{-7}.Lp(14) = p(14)\,L^6p(14) = p(14)p(45)p(14) = p(15),$$

and hence

$$P^{-1}RQ = p(15)\,L^6 = p(15)p(45) = p(45)p(14) = L^6p(14) = Qp(14).$$

Thus $QP^{-1}RQ = p(14)$ and so $p(14)$ and $Rp(14) = L$ belong to the group generated by P and Q. Thus, since L and $p(14)$ generate G, so also do P and Q.

A closely related pair of generators has been given by the author (Todd 1936, p. 173), namely, a pair of operations S and T each of period 5, such that ST is of period 9 and ST^{-1} is of period 3. Such a pair of operations is given in fact by

$$S = QP, \quad T = PQ,$$

where P and Q are the operations just described. For these are of period 5, and $ST = QP^2Q$ is conjugate to P^2 (and so of period 9), while

$$ST^{-1} = QPQ^{-1}P^{-1} = Lp(14)p(45)p(14)\,L^{-7} = L[p(14)p(45)]^2\,L^{-1},$$

which is of period 3. And, since $Q^2 = 1$, $TS = P^2$ and $(TS)^5 = P$. Thus

$$P = (TS)^5, \quad Q = (TS)^{-5}\,T,$$

are expressible in terms of S and T. Whence, as P and Q generate G, so also do S and T.

It seems rather striking that the pairs of generators found, in very different ways, by various writers, should prove to be related to each other in so simple a manner.

I should like to express my thanks to Professor Baker, who has very kindly placed the proof sheets of his tract at my disposal, and with whom I have had many stimulating discussions on the subject of this paper.

REFERENCES

Baker 1946 *A locus with* 25920 *linear self-transformations*. Camb. Univ. Press.
Brahana 1930 *Ann. Math.* (2), **31**, 529–549.
Burkhardt 1890 *Math. Ann.* **38**, 161–224.
Burnside 1911 *Theory of groups of finite order*. 2nd ed. Camb. Univ. Press.
Coble 1906 *Amer. J. Math.* **28**, 333–366.
Coxeter 1940 *Amer. J. Math.* **62**, 457–486.
Dickson 1901 *Linear groups, with an exposition of the Galois field theory*. Leipzig: Teubner.
Frame 1936 *Duke Math. J.* **2**, 477–484.
Jordan 1870 *Traité des substitutions*. Paris: Gauthier-Villars
Moore 1897 *Proc. Lond. Math. Soc.* (1), **28**, 357–366.
Todd 1936 *Quart. J. Math.* (Oxford series), **7**, 168–174.

The inversion spectrum of ammonia at centimetre wave-lengths

BY B. BLEANEY AND R. P. PENROSE, *Clarendon Laboratory, Oxford*

(*Communicated by Lord Cherwell, F.R.S.—Received* 17 *July* 1946—
Read 27 *February* 1947)

A technique is described for the investigation of gaseous spectra at high radio-frequencies. The sensitivity of the method is such that an absorption of 2×10^{-6} per cm. of path can be detected at wave-lengths around 1 cm.

The inversion spectrum of ammonia in the ground vibrational state has been examined at wave-lengths between 1·1 and 1·6 cm. and found to contain an extensive fine structure, arising from centrifugal distortion of the molecule. Twenty-nine lines have been identified, each corresponding to a different rotational quantum state. The wave numbers of these lines can be accurately represented by the formula

$$\bar{\nu}=0{\cdot}7935+[-0{\cdot}0050_5\overline{J^2+J}+0{\cdot}0070_4K^2]+0{\cdot}63[-0{\cdot}0050\overline{J^2+J}+0{\cdot}0070K^2]^2\ \text{cm.}^{-1}\ \text{(vacuo)}$$

where J, K are the rotational quantum numbers of the symmetrical top molecule.

Detailed measurements of 18 lines at a pressure of 0·5 mm. Hg show that their half widths, due to collision broadening, lie between 2 and 5×10^{-4} cm.$^{-1}$, while their intensities agree with the theoretical values, within the experimental error of ±5 %. The integrated intensity at a pressure of 4·5 mm. Hg, over the interval 0·67 to 0·87 cm.$^{-1}$, agrees with the calculated intensity to better than 1 %.

1. INTRODUCTION

The well-known doubling of most of the lines of the infra-red spectrum of gaseous ammonia is associated with the existence of two equilibrium configurations of the molecule, with the nitrogen atom on either side of the plane containing the three hydrogen atoms. The appropriate potential energy curve of the molecule has two identical minima, separated by a fairly low hump; the associated 'resonance' causes a splitting of the energy levels ('inversion doubling'). In the ground vibrational state this splitting amounts to 0·67 cm.$^{-1}$, determined by Wright & Randall (1933) from measurements on the pure rotation spectrum in the far infra-red. Transitions between the two sub-levels are allowed, and an absorption line would be expected at a wave-length of about 1·5 cm. The existence of this absorption was verified experimentally by Cleeton & Williams (1934) using split-anode magnetrons as sources of radiation. The wave-length was determined by a diffraction grating, and the absorption was measured by inserting ammonia (at atmospheric pressure in a cloth bag) in the path of radiation. The peak of the absorption curve was found to be at about 0·8 cm.$^{-1}$, and the absorption was spread over a considerable frequency range (0·5 cm.$^{-1}$) owing to collision broadening.

This absorption has been re-examined using wave-guide technique. Low gas pressures, down to 0·2 mm. Hg, have been employed in order to reduce the effects of collision broadening, and an extensive fine structure of the absorption has thereby been revealed. This fine structure is due to centrifugal distortion of the molecule, the effect of which is comparatively large because the inversion frequency varies very rapidly with the distance between the nitrogen atom and the plane of the hydrogen atoms.

The absorption coefficients to be expected at centimetre wave-lengths are much smaller than those encountered in the infra-red, and direct measurement would require considerable path lengths. The difficulties inherent in such a measurement have been avoided by the use of a resonant cavity, which, effectively, is a short length of wave-guide traversed by the radiation many times before its final decay. The efficiency of such a cavity is measured by its magnification or Q value, defined as 2π times the ratio of the stored energy to the energy dissipated per cycle. On admitting an absorbing gas to the resonator its Q value is reduced, say, from Q_0 to Q_1. The absorption coefficient of the gas is related to Q_0, Q_1 by the expression

$$\gamma = \frac{2\pi}{\lambda}\left(\frac{1}{Q_1} - \frac{1}{Q_0}\right). \tag{1}$$

The Q values were determined from the resonance curve obtained by detuning the cavity, and were subject to errors of a few per cent, owing to fluctuations of the frequency of the oscillator. Since the gaseous absorption is found from the difference of the two values of $1/Q$ the inaccuracy in its determination is considerable unless $1/Q_1$ is appreciably different from $1/Q_0$. This method is thus limited to absorptions comparable with the 'equivalent absorption' $2\pi/\lambda Q_0$ of the cavity. The value of Q_0 was about 10,000, from which $2\pi/\lambda Q_0 \approx 5 \times 10^{-4}$ per cm.

To extend the method to smaller absorptions, use was made of the fact that the power output from the oscillator was constant within less than 1 %. When the cavity is tuned to resonance, small fluctuations in the oscillator frequency have only a second order effect on the amplitude of the energy density in the cavity, instead of the first order effect obtained when the resonator is detuned for the measurement of Q. The energy density in the cavity can be determined by a loosely coupled detector, being proportional to the power registered by this detector, a crystal rectifier. Simple equivalent circuit considerations show that this power δ varies as Q^2, giving

$$\frac{Q_1}{Q_0} = \sqrt{\frac{\delta_1}{\delta_0}}. \tag{2}$$

Hence we have $$\gamma = \frac{2\pi}{\lambda Q_1}\left(1 - \sqrt{\frac{\delta_1}{\delta_0}}\right) = \frac{2\pi}{\lambda Q_0}\left(\sqrt{\frac{\delta_0}{\delta_1}} - 1\right). \tag{3}$$

Thus it is possible to determine the absorption by measuring *either* Q_1 *or* Q_0 together with the power ratio δ_1/δ_0. The stability of the oscillator (a reflexion klystron) was such that a difference between δ_0 and δ_1 of 1 % could be detected, corresponding to an absorption coefficient of about 2×10^{-6} per cm. At the higher attenuations, the values of γ given by (1) and (3) were compared, and the agreement was generally satisfactory. The probable error of the mean values of the absorption coefficients obtained from this comparison is estimated at about ± 3 % or $\pm 2 \times 10^{-6}$ per cm., whichever is the greater.

The wave-length of the radiation was measured by a second resonant cavity. It was carefully designed to give an accuracy of 2 parts in 10,000 or better. Small changes in wave-length, such as were used in measuring the breadths of lines at low pressure, could be measured more accurately, the limit of reading corresponding to a change in wave-number of about 0·00002 cm.$^{-1}$. This wave-meter was used because apparatus for the measurement of frequency by means of the harmonics of quartz crystals was not available. The latter method would give greater accuracy, and would also be desirable because the frequency of a line is more fundamental than the wave-length, through its relation to the quantum of energy.

The essential differences between the technique described in this paper and that normally used in the infra-red may be summarized as follows:

(*a*) The radiation from the source is collimated by a wave-guide, which produces a beam of radiation less than a wave-length across. This avoids the use of mirrors, which would have to be several metres in diameter to collimate the radiation at these wave-lengths.

(*b*) The effective multiplication of path length obtained by using a resonator makes it possible to detect absorptions of the order of 10^{-6} per cm. The volume of the resonator is less than 10 c.c. and the gas can be rapidly admitted or pumped away. The energy densities in the cavity with and without gas could therefore accurately be compared without requiring long-period stability in the oscillator.

(*c*) The reflexion klystron is virtually a monochromatic source, and the limit of resolution is determined by the widths of the absorption lines themselves, at the gas pressures used in these experiments. In the infra-red the sources provide a continuous spectrum of radiation and the resolving power is that of the spectrometer.

Although the methods described in this paper for the examination of gaseous spectra have been applied only over a small wave-length interval between 1 and 1·5 cm., they are capable of extension to decimetre, centimetre and millimetre wave-lengths. Oscillators and detectors are available at all except millimetre wave-lengths, and it is likely that they will be developed for this region soon, as work at 5 mm. has already been reported. The strength of the absorptions to be expected, and the sensitivity of the apparatus, vary with roughly similar powers of the wave-length. *Ceteris paribus*, the absorption coefficients vary inversely as the square of the wave-length (see § 5, equation (4)). If the resonant cavity is scaled in proportion to the wave-length, its Q varies with $\sqrt{\lambda}$, and the smallest detectable absorption will, by equation (3), vary with $\lambda^{-\frac{3}{2}}$. Thus for measurements at wave-lengths between 100 and 1 mm., the effective merit of the method changes by a factor of 10. At wave-lengths beyond this interval, the limitations of the method are mainly mechanical; at decimetre wave-lengths the cavity becomes inconveniently bulky, while its increasing Q requires still greater frequency stability of the oscillator. At the shorter wave-lengths the minute dimensions cause machining difficulties, and the use of a large cavity resonant in many modes, instead of a small cavity resonant in one particular mode (see Autler, Becker & Kellogg 1946) appears more promising.

2. APPARATUS

Resonator

A cylindrical cavity was used, shown in figure 1. It was tuned to resonance by moving the plunger A by means of a micrometer head. The H_0 mode of resonance was employed because it has the highest Q value for a given size of cavity; a special system of coupling between cavity and wave-guide avoided the excitation of other modes as far as possible. This coupling system is described by the authors elsewhere (1946 *a*). Further discrimination against unwanted modes was provided by a vacuum-tight glass tube (B), containing water, attached to the back of the plunger. All modes except the H_0 tend to leak out of the main cavity past the plunger, and are damped by the strong absorption of the water.

To make the cavity vacuum tight, thin mica sheets were waxed over the coupling holes on the outside of the resonator. The problem of moving the plunger inside the vacuum was overcome by sealing the annular space (C) between the micrometer spindle and its bushing with a viscous oil of low vapour pressure. By this means the position of the plunger could be controlled with the full accuracy of the micrometer screw; the thrust of the atmosphere on the spindle was sufficient to eliminate back-lash. Other joints in the resonator were waxed over or sealed with oil; the pressure inside could be reduced below 10^{-3} mm. Hg by a mercury diffusion pump.

Gas was admitted to the chamber through the hole (D) outside the high-frequency field. The higher pressures were measured on a mercury manometer; a McLeod gauge was used for those below 2 mm.

The plunger travel was sufficient to give at least the first two H_0 resonances at all the frequencies employed. The plunger movement needed to tune between the half-power points was about 0·005 mm.; the corresponding micrometer head rotation was measured by a lamp and scale method which gave an over-all magnification of the plunger movement of 12,000.

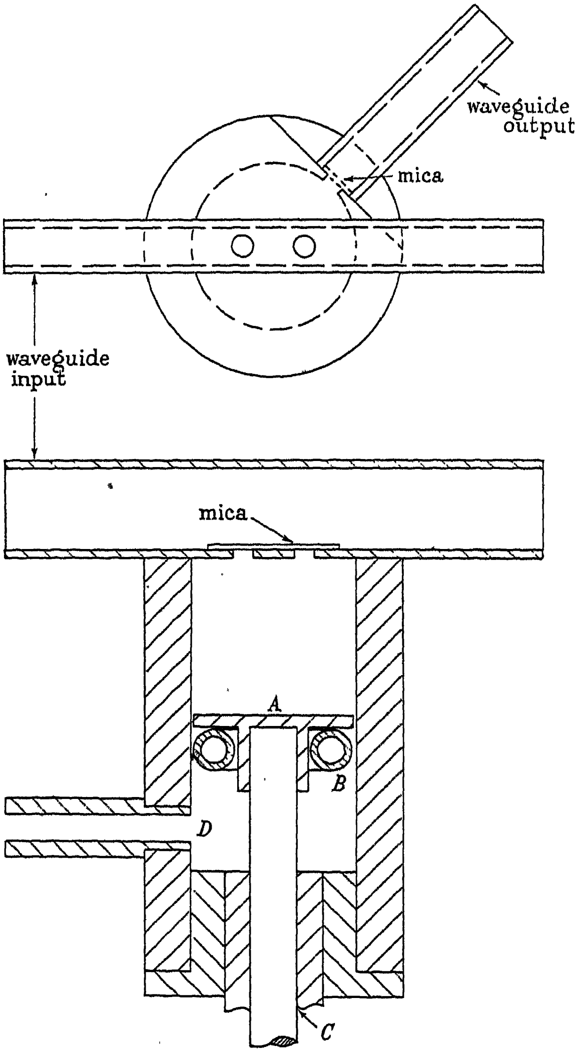


FIGURE 1. Resonator.

Wave-meter

The H_0 cavity used for wave-length measurement was coupled in the same way as the experimental cavity, but it was larger in diameter. This makes the readings less sensitive to machining errors in the barrel. Inaccuracies in the barrel diameter and in the micrometer head made roughly equal contributions to the over-all error in the wave-length value; this was believed accurate to 1 part in 5000.

Oscillator and detector

The power source was a reflexion klystron designed and constructed in this laboratory. The power available on the output side of the resonator was so small that a crystal detector had to be used. The accuracy with which attenuation can be measured depends on the extent to which the detector output is proportional to the incident power. The response characteristics of several crystals were compared with that of a thermocouple, whose reading should give a true indication of power. These experiments indicated that the error introduced in the attenuation values by deviations of the crystals from the square law should be less than 5 %.

An attenuator in the resonator output feeder served to adjust the power at the detector to a convenient value.

3. EXPERIMENTAL

At each wave-length the Q value of the resonator was measured by determining the movement of the plunger required to tune the resonator between the half-power points. With the optical magnification this corresponded to deflexions of between 5 and 20 cm. and the average of a number of readings was taken. The scatter in these readings was usually a few per cent, caused mainly by fluctuations in the frequency of the oscillator. A check on the accuracy of this determination was obtained by admitting to the resonator a quantity of ammonia sufficient to diminish the Q appreciably. The new value of Q was measured, together with the power fed through the cavity at resonance before and after admitting the ammonia. The ratio of the square root of these powers should equal the ratio of the two values of Q (equation (2) above). In general the difference between the two ratios was within the experimental error, and an average value of the change in Q was used to calculate the absorption. Measurements of the absorption coefficients at a number of pressures were made, the smaller absorptions being determined from the power change registered by the crystal detector. The absorption coefficients at the various standard pressures (e.g. 4·5, 1·2, 0·5 mm. Hg) were obtained by interpolation on a graph of absorption coefficient against pressure.

Measurements at small frequency intervals were made to delineate each absorption line at pressures of 1·2 and 0·5 mm. Hg. At the latter pressure most of the lines were resolved well enough to make possible an accurate determination of their breadths. A few lines were separated only at a still lower pressure (0·2 mm.).

Whenever possible, the exact centre of each line was found by a brief examination of the absorption at a pressure of less than 1/10 mm. The value thus determined was in close agreement with that deduced from the absorption curve for 0·5 mm. pressure.

4. THE SPECTRUM

The absorption curve obtained at a pressure of 1·2 mm. Hg is shown in figure 2. The measurements cover the wave-number band from 0·65 to 0·87 cm.$^{-1}$, which includes all the more intense lines. The coordinates are given both in terms of absorption coefficient per centimetre path (the usual spectroscopic unit) and of decibels per kilometre. The latter unit is customary in radio engineering and is particularly convenient for the small absorptions to be expected at these wavelenths. The conversion factor is given by

$$1 \text{ db./km.} = 10^6(\log_{10} e) \times \text{absorption coefficient/cm.}$$

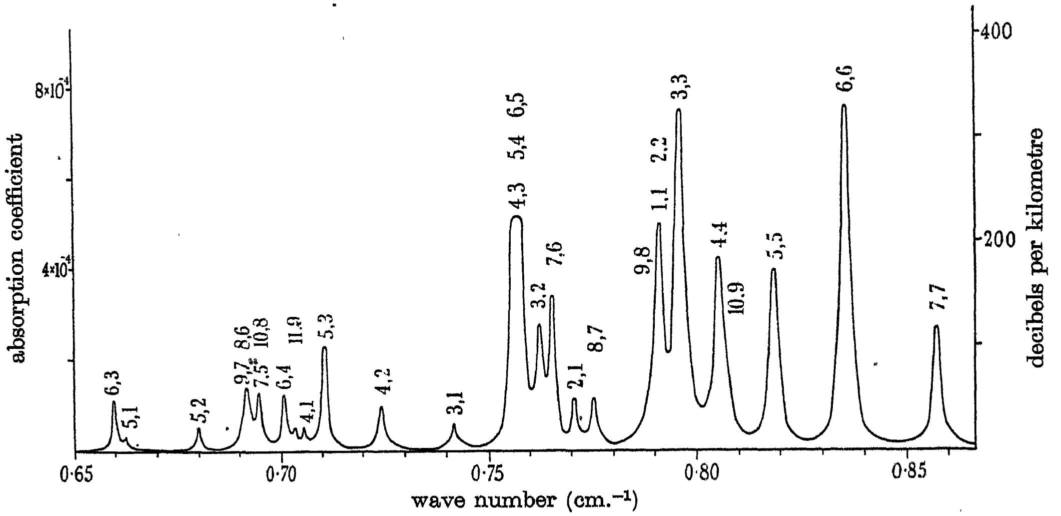


FIGURE 2. The spectrum of ammonia near 0·8 cm.$^{-1}$.

The widths of the lines shown in figure 2 are about 10^{-3} cm.$^{-1}$. The widths to be expected from radiation damping ('natural line breadth') and the Doppler effect are respectively some 10^{-17} and 10^{-6} cm.$^{-1}$. Thus, even at a pressure of 1·2 mm. Hg, only collision broadening of the line need be considered, and the lines would be expected to narrow still further at lower pressures. This has been confirmed using pressures down to 1/10 mm. Hg. The actual breadths of the lines indicate a collision diameter several times greater than the kinetic theory value. This may be attributed to the strong electric field of the ammonia dipole, and will be discussed in another paper.

Identification of the lines

The theory of the inversion splitting of the energy levels of the ammonia molecule has been considered by Dennison & Uhlenbeck (1932). Using the Wentzel-Kramers-Brillouin approximation, they showed that the ratio of the splitting Δ to the normal vibration frequency $\bar{\nu}$ should vary exponentially with the area cut from the potential hump by the vibrational energy level, being thus a delicate function of the height of the pyramid. Only in the case of a very flat pyramidal molecule, such as ammonia, is the splitting in the ground vibrational state large enough to be detected, and in this case the slight centrifugal distortion of the molecule in the higher rotational energy levels may cause a considerable change in the splitting, and therefore in the inversion frequency.

The effect of centrifugal distortion can be seen qualitatively from the following considerations. When the molecule is rotating mainly about its symmetry axis, the hydrogen atoms tend to fly outwards, thus decreasing the height of the pyramid and increasing the inversion frequency. Rotation about an axis perpendicular to the symmetry axis will have the opposite effect. The energies associated with these two rotations depend on K^2 and (J^2+J-K^2) respectively, and, by an extension of Dennison & Uhlenbeck's treatment, Hsi-Yin Sheng, Barker & Dennison (1941) have shown that $\bar{\nu}$ should satisfy the relation

$$\begin{aligned}\bar{\nu} &= \bar{\nu}_0 - A'(J^2+J-K^2)+B'K^2\\ &= \bar{\nu}_0 - A(J^2+J)+BK^2.\end{aligned}$$

They have examined the fundamental vibrational band at 10μ and have shown that their observations can be represented by a formula of this kind. They have also calculated the constants A and B, using the potential energy curve proposed by Manning (1935) and obtain values in satisfactory agreement with the experimental determinations. By a similar computation, they find for the ground level of the vibration

$$\bar{\nu} = \bar{\nu}_0 - 0\cdot0011(J^2+J)+0\cdot0016K^2\,\text{cm.}^{-1}.$$

In attempting to find a formula of this kind which would adequately represent the system of lines of figure 2, it was assumed that the series of strong lines at the high frequency end of the spectrum corresponded to the levels $J=K$ which have both large statistical weights and large transition probabilities.

The fact that the levels for which K is a multiple of three have twice the *a priori* statistical weight (due to nuclear spin) of the others suggested that the lines at 0·797 and 0·836 cm.$^{-1}$ must be identified as $(J, K) = (3, 3)$ and $(6; 6)$. This yielded the values

$$A = 0\cdot005_0\,\text{cm.}^{-1},\quad B = 0\cdot007_0\,\text{cm.}^{-1},\quad \bar{\nu}_0 = 0\cdot794\,\text{cm.}^{-1}.$$

With these constants the positions of the remaining lines could be predicted, and fair agreement was found with the frequencies of the lines found experimentally. The discrepancies between the predicted and observed values increased steadily as the observations advanced towards the lower frequencies. This was not surprising, for $(\bar{\nu}-\bar{\nu}_0)$ amounted to 20 % of $\bar{\nu}_0$ at the lower limit of the frequency band. In the calculation of Hsi-Yin Sheng, Barker & Dennison the formula for the splitting includes only the first order term in the expansion of an exponential, and other effects which would yield second-order terms are omitted. In the ground vibrational state the fractional change in the splitting is so large that the second-order term from the exponential will be considerable, and it can be included without the introduction of further unknown constants. The formula*

$$\begin{aligned}\bar{\nu} = 0\cdot7935 - 0\cdot0050_5(J^2+J)+0\cdot0070_4K^2&\\ +0\cdot63\{-0\cdot0050(J^2+J)+0\cdot0070K^2\}^2\,&\text{cm.}^{-1}\ (\text{vacuo})\end{aligned}$$

---

* The value for $\bar{\nu}_0$ given here (0·7935 cm.$^{-1}$) differs slightly from that in a preliminary report (Bleaney & Penrose 1946 *b*) where, owing to an error in correcting the wave numbers to vacuum, a value of 0·7940 cm.$^{-1}$ was given.

was adopted as giving the best fit with the experimental results. In table 1 the observed frequencies of the lines are given, together with the values (in italics) calculated from this formula.

The calculated frequencies for the $K = 0$ states are included for comparison with the value of the inversion splitting found by Wright & Randall in their experiments on the rotational spectrum (vide infra).

TABLE 1. WAVE-NUMBERS OF LINES IN THE SPECTRUM

Figures in italics are calculated from the formula

$$\bar{\nu} = 0\cdot7935 - 0\cdot0050_5(J^2+J) + 0\cdot0070_4K^2 + 0\cdot63\{-0\cdot0050(J^2+J) + 0\cdot0070K^2\}^2 \text{ cm.}^{-1} \text{ (vacuo)}$$

J \ K	0	1	2	3	4	5	6	7	8	9
0	(*0·7935*)									
1		0·7904								
	(*0·7834*)	*0·7904*								
2		0·7706	0·7913							
	(*0·7638*)	*0·7705*	*0·7913*							
3		0·7418	0·7617	0·7964						
	(*0·7352*)	*0·7417*	*0·7617*	*0·7963*						
4		0·7050	0·7240	0·7569	0·8052					
	(*0·6989*)	*0·7050*	*0·7239*	*0·7568*	*0·8052*					
5		0·6616	0·6795	0·7100	0·7556	0·8183				
	(*0·6564*)	*0·6624*	*0·6797*	*0·7103*	*0·7555*	*0·8184*				
6				0·6590	0·7003	0·7583	0·8360			
	(*0·6098*)	*0·6149*	*0·6308*	*0·6587*	*0·7002*	*0·7582*	*0·8359*			
7						0·6939	0·7648	0·8581		
						0·6938	*0·7646*	*0·8582*		
8							0·6910	0·7750		
							0·6909	*0·7751*		
9								0·6916	0·7892	
								0·6915	*0·7895*	
10									0·6954	0·8076
									0·6954	*0·8083*
11										0·7031
										0·7029

The agreement between the experimental and calculated values is very good. The greatest discrepancy (0·0008 cm.$^{-1}$) occurs for the line (5, 1), for which even the third order term in the exponential would be appreciable. The agreement leaves little doubt as to the correctness of the identification of the lines, which is confirmed by comparison of the observed and theoretical intensities (see below).

Since the work described in this paper was completed another investigation of the ammonia spectrum has been briefly reported by Good (1946*a*). The narrow lines obtained at a pressure of 1/10 mm. Hg or less were displayed on a cathode ray tube screen. The frequencies of the individual lines are not given, except in one case, 23,878 Mcyc./sec. This is presumably the line (3, 3) and is excellent agreement with our value of 0·7964 ± 0·0001 cm.$^{-1}$, which corresponds to 23,876 ± 3 Mcyc./sec. The formula suggested to represent the wave-numbers of the lines is

$$\bar{\nu} = 0{\cdot}7932 - 0{\cdot}0048(J^2 + J - K^2) + 0{\cdot}0020K^2.^*$$

The difference between the coefficients in this expression and the corresponding coefficients in ours may be due to adjustment of the values to give the best fit on all 28 lines observed without the use of a second-order term. This would lead to the choice of a smaller numerical value for the negative coefficient of $(J^2 + J - K^2)$, since the second-order term is essentially positive. There is no doubt that the identification of the lines is the same as made by us.

The experimental values of the constants A, B of the splitting are considerably greater than those calculated by Sheng, Barker & Dennison. Their calculation rests on the potential energy curve for the ammonia molecule proposed by Manning (1935). This curve has no firm foundation, and was adopted to obtain an exact solution of the wave-equation, the constants being adjusted to fit the energy levels with the observed vibrational frequencies. It gives satisfactory values for the inversion splitting of the vibrational levels, and for the fine structure of the 10μ vibrational band. Modification of this potential curve, to give better correlation with the results reported in this paper, would presumably be confined to the region where the potential energy is not much greater than the energy of the ground vibrational level. This would not affect the agreement with the previous observations.

5. MAGNITUDE OF THE ABSORPTION

Since the experimental errors in the measurement of the absorption coefficients are not expected to amount to more than a few per cent, an accurate comparison with the values predicted by the quantum mechanical intensity formula is possible. This formula is (cf. Van Vleck & Weisskopf, 1945)

$$\int \frac{\gamma}{\nu^2}\,d\nu = \frac{8\pi^3 N_{JK}}{3ckT}\,|\mu_{JK}|^2, \tag{4}$$

since $h\nu \ll kT$ in this region. Here N_{JK} is the number of molecules per c.c. occupying the level with rotational quantum numbers (J, K) and $|\mu_{JK}|^2$ is the transition probability for the inversion. Dennison (1931) has shown that

$$|\mu_{JK}|^2 = \frac{\mu^2 K^2}{J^2 + J},$$

---

* *Note added in proof.* In a later paper Good (1946*b*) gives the formula

$$\bar{\nu} = 0{\cdot}79347 - 0{\cdot}005048(J^2 + J) + 0{\cdot}007040K^2 + 0{\cdot}00001546(J^2 + J)^2 - 0{\cdot}00004260(J^2 + J)K^2 + 0{\cdot}00002920K^4,$$

where all the constants are empirical. The close agreement between this formula and that proposed by us, which involves only three empirical constants, has been pointed out elsewhere (Bleaney & Penrose 1946*c*). The accuracy of determination of the wave-numbers of the lines claimed by Good (± 0·02 %) is the same as in our experiments.

where μ is the permanent dipole moment of ammonia, for which the value $1{\cdot}44 \times 10^{-18}$ e.s.u. is taken in these calculations.

For comparison with the experimental results two methods have been used. The first of these involves a determination of the total absorption due to a number of lines; from equation (4) it is obvious that this requires evaluation of the area contained under a curve of (γ/ν^2) against ν. This area is then to be compared with the theoretical value obtained by summation of the appropriate values of $(N_{JK} \mid \mu_{JK} \mid^2)$ for the lines whose wave-numbers lie in the range of integration. In computing the area under the experimental curve greater accuracy is obtained by using a curve determined at a pressure where the individual lines are smoothed over by the pressure broadening. The pressure should not be too high, however, for it is desirable to keep most of the absorption within the band of frequencies covered by the experiments.

The absorption curve measured at 4·5 mm. pressure was chosen as the most suitable and the area between $\bar{\nu} = 0{\cdot}67$ and $0{\cdot}87$ cm.$^{-1}$ was found to be

$$\int \frac{\gamma}{\bar{\nu}^2} d\bar{\nu} = 44{\cdot}5 \text{ db./km.cm.}$$

A simple calculation indicated that this figure should be increased by 4 % for the tails at the high and low ends of the frequency band, giving 46·3 db./km.cm. This figure includes, of course, the tails of those lines whose centres lie outside the band, but their contribution is less than 1 % and can be neglected. The value calculated from the statistical weights and transition probabilities is $\int \frac{\gamma}{\bar{\nu}^2} d\bar{\nu} = 46{\cdot}5$ db./km.cm. The exact agreement of this and the experimental value is somewhat fortuitous, as a difference of several per cent would not lie outside the experimental errors. The latter are however mainly random in nature and the process of computing the integral has a similar effect to taking the mean of a large number of observations.

A second comparison between the theoretical and experimental absorption coefficients can be obtained from the intensities of individual lines. This requires that the lines be sufficiently resolved to make possible a measurement of the total area under each line separately, or, what is equivalent, measurement of the breadths of the lines at half intensity. From the latter and the absorption at the centre of the line the total intensity can be found, assuming the shape of the absorption curve to be given by the usual simple formula

$$\frac{\gamma}{\bar{\nu}^2} \propto \frac{\Delta\bar{\nu}}{(\bar{\nu}-\bar{\nu}_0)^2 + \Delta\bar{\nu}^2}.$$

In this formula $\Delta\bar{\nu}$ is half the breadth of the line at half-intensity, i.e. half of what is usually termed the 'breadth' in spectroscopy. We shall refer to it as the 'line-breadth constant'.

To obtain rather better resolution of the lines than at 1·2 mm. pressure (figure 2), a pressure of 0·5 mm. was chosen at which to make a detailed examination of each line. At this pressure all but a few lines were well resolved; these few were examined at lower pressures (e.g. 0·2 mm.) to determine their exact frequencies, though not their breadths. A typical absorption curve is shown in figure 3, the readings of the micrometer head being plotted as abscissae. The limit of reading of the micrometer head corresponds to a frequency change of about 0·00002 cm.$^{-1}$; thus small changes of wave-number can be measured with a probable error of this order. The micrometer reading is linearly related to the length l of the cavity, and changes in the reading can be reduced to changes in wave-number by the relation

$$\frac{\Delta\bar{\nu}}{\Delta\lambda_g} = \frac{n\Delta\bar{\nu}}{\Delta l} = -\frac{1}{\bar{\nu}\lambda_g^3},$$

where λ_g is the wave-length in the cavity, determined from the positions of successive resonances, and n is the number of full wave-lengths in the cavity at the particular resonance used for the examination of the line. Both $\bar{\nu}$ and λ_g and, therefore, the conversion factor are known to a few parts in 10^4.

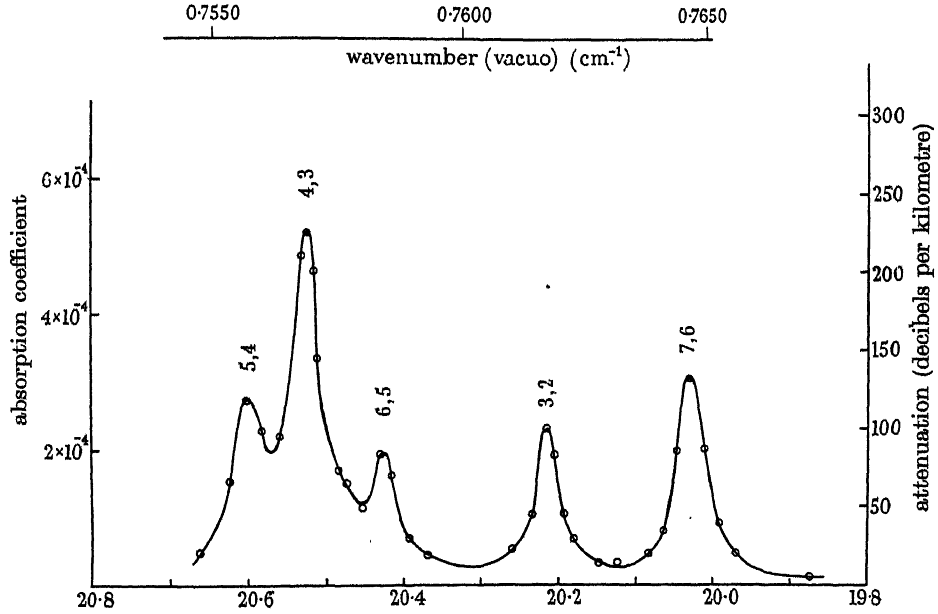


FIGURE 3. Detail of spectrum at lower pressure (0·5 mm.) showing improved resolution.

In table 2 the experimental values of the absorption coefficient (γ_0) at the centre of the line, and the line-breadth constant ($\Delta\bar{\nu}$), are given for some seventeen lines (columns 2 and 3). In the fourth column is shown the intensity $I'_{JK}/\bar{\nu}^2$ calculated from the formula

$$\frac{I'_{JK}}{\bar{\nu}^2} = \frac{\pi\gamma_0}{\bar{\nu}^2}\Delta\bar{\nu}(10^{-5}V_0) \quad (V_0 = \text{molar volume at 0·50 mm. pressure}),$$

where the constant $10^{-5}V_0$ represents a reduction from a kilometre path to a path length such that a mole of gas would be contained per sq.cm. of cross-section. The units of $I'_{JK}/\bar{\nu}^2$ are therefore decibels per mole. The corresponding values of $I'_{JK}/\bar{\nu}^2$ calculated from the theoretical formula (4) are given in column 5, and do not deviate from the experimental values by more than ± 5 %. The measurements of the line-breadth constants, which range from 2×10^{-4} to 5×10^{-4} cm.$^{-1}$, are liable to errors of the order ± 5 %. To compare the averages, the values of columns 4 and 5 have been summed separately; they agree very closely showing that there is no consistent deviation.

TABLE 2. COMPARISON OF MEASURED AND CALCULATED INTENSITIES

line (J, K)	$\Delta\bar{\nu}$ at $p = 0{\cdot}5$ mm. $\times 10^{-4}$ cm.$^{-1}$	γ_0 (db./km.)	$\frac{I'_{JK}}{\bar{\nu}^2} = \frac{\pi\gamma_0}{\bar{\nu}^2}\Delta\bar{\nu}(10^{-5}V_0)$ (db./mol.)	$\frac{I_{JK}}{\bar{\nu}^2}$ (calc.)
2, 1	2·6	51	25	27
3, 1	2·4	29	14·5	14
3, 2	3·2	100	62·5	59
3, 3	4·5	360	289	289
4, 4	4·5	190	149	151
5, 1	1·8	7	3·3	3·5
5, 2	2·6	20	13	15
5, 3	3·3	96	71	73·5
5, 5	4·7	186	148	137
6, 3	3·1	44	35·5	34
6, 4	3·6	42	35	34·5
6, 6	4·7	276	198	210
7, 5	4·1	28	27	27
7, 6	3·9	132	100	95·5
8, 7	4·1	44	34	32·5
10, 9	4·2	35	25·5	23
11, 9	2·9	8·5	5·7	6·4
		Total	1237	1236

The good agreement between the theoretical and experimental values of the absorption coefficients by both methods of comparison is in marked contrast with results obtained in the infra-red, where the experimental difficulties generally preclude accurate measurement of intensity. In the case of ammonia these difficulties were largely overcome by Foley & Randall (1941); from the intensity of the line at 270 cm.$^{-1}$ they deduced a value of 1·30 Debye units for the permanent dipole moment. If the measurements reported in this paper are regarded as yielding a dipole moment, the value 1·44 ± 0·01 Debye units is obtained. The values obtained by the usual methods are 1·44 (Zahn 1926), 1·48 (Watson 1927), 1·44 (Keyes & Kirkwood 1930) and 1·437 (Van Itterbeek & De Clippeleier 1946).

6. COMPARISON WITH WRIGHT AND RANDALL

The most intense lines in the spectrum described in this paper lie in the neighbourhood of 0·8 cm.$^{-1}$. On the other hand, Wright & Randall (1933) found that the splitting of the 4th, 5th and 6th rotational lines in the far infra-red (which should be twice the inversion frequency) was 1·33 cm.$^{-1}$. There is no real discrepancy between these two values, however, because the transition probabilities are different for rotational lines ($\Delta J = \pm 1$) and the lines in the long-wave spectrum for which ($\Delta J = 0$). In fact the lines corresponding to $K = 0$, which are forbidden in the inversion spectrum, are those with the greatest intensity in the rotation spectrum. The inversion frequencies (table 2), observed or predicted by the splitting formula, have been used to calculate the mean splitting of the rotational transition $J = 5$ to $J = 6$, the lines corresponding to different values of K being weighted according to their statistical weights and transition probabilities. The value thus obtained was 1·35 cm.$^{-1}$ in good agreement with that observed by Wright & Randall. Only the mean values can be compared, since there is an additional splitting of the K sub-levels due to centrifugal distortion which must also be considered in reconstructing the shape of the unresolved lines observed by Wright & Randall.

We wish to make acknowledgement to Lord Cherwell for extending to us the facilities of his laboratory; and to Mr K. W. H. Stevens for his kind help with the experiments. The work was conducted under the auspices of the Board of Admiralty, to whom we are indebted for permission to publish this paper.

REFERENCES

Autler, Becker & Kellogg 1946 *Phys. Rev.* **69**, 694.
Bleaney & Penrose 1946*a* *Proc. Phys. Soc.* (awaiting publication).
Bleaney & Penrose 1946*b* *Nature*, **157**, 339.
Bleaney & Penrose 1946*c* *Phys. Rev.* **70**, 775.
Cleeton & Williams 1934 *Phys. Rev.* **45**, 234.
Dennison 1931 *Rev. Mod. Phys.* **3**, 280.
Dennison & Uhlenbeck 1932 *Phys. Rev.* **41**, 313.
Foley & Randall 1941 *Phys. Rev.* **59**, 171.
Good 1946*a* *Phys. Rev.* **69**, 539.
Good 1946*b* *Phys. Rev.* **70**, 213.
Hsi-Yin Sheng, Barker & Dennison 1941 *Phys. Rev.* **60**, 786.
Keyes & Kirkwood 1930 *Phys. Rev.* **36**, 1570.
Manning 1935 *J. Chem. Phys.* **3**, 136.
Van Itterbeek & De Clippeleier 1946 *Physica*, **12**, 97.
Van Vleck & Weisskopf 1945 *Rev. Mod. Phys.* **17**, 227.
Watson 1927 *Proc. Roy. Soc.* A, **117**, 43.
Wright & Randall 1933 *Phys. Rev.* **44**, 391.
Zahn 1926 *Phys. Rev.* **27**, 455.

Infinite irreducible representations of the Lorentz group

By Harish-Chandra, *Gonville and Caius College, Cambridge*

(*Communicated by P. A. M. Dirac, F.R.S.—Received* 8 *August* 1946)

It is shown that corresponding to every pair of complex numbers κ, κ^* for which $2(\kappa-\kappa^*)$ is real and integral, there exists, in general, one irreducible representation $\mathfrak{D}_{\kappa,\kappa^*}$ of the Lorentz group. However, if 4κ, $4\kappa^*$ are both real and integral there are two representations $\mathfrak{D}^+_{\kappa,\kappa^*}$ and $\mathfrak{D}^-_{\kappa,\kappa^*}$ associated to the pair (κ,κ^*). All these representations are infinite except $\mathfrak{D}^-_{\kappa,\kappa^*}$ which is finite if 2κ, $2\kappa^*$ are both integral. For suitable values of (κ,κ^*), $\mathfrak{D}_{\kappa,\kappa^*}$ or $\mathfrak{D}^+_{\kappa,\kappa^*}$ is unitary. $\mathfrak{U}$ and $\mathfrak{B}$ matrices similar to those given by Dirac (1936) and Fierz (1939) are introduced for these infinite representations. The extension of Dirac's expansor formalism to cover half-integral spins is given. These new quantities, which are called expinors, bear the same relation to spinors as Dirac's expansors to tensors. It is shown that they can be used to describe the spin properties of a particle in accordance with the principles of quantum mechanics.

1. INTRODUCTION

All the finite irreducible representations of the Lorentz group are well known. Every such representation is characterized by an ordered pair of numbers p, q such that $2p$ and $2q$ are integral and $\geqslant 0$. None of these representations, however, is unitary. Dirac (1945) has recently drawn attention to the existence of some unitary, though infinite, representations with a view to their possible physical applications. The present paper is concerned with the investigation of the general irreducible representations of the proper Lorentz group. It is found that such a representation can, in general, be characterized by an ordered pair (κ,κ^*) of complex numbers such that $2(\kappa-\kappa^*)$ is an integer† and may therefore be denoted by $\mathfrak{D}_{\kappa,\kappa^*}$. However, in case both κ and κ^* are of the form $\frac{1}{4}n(n\neq-2)$ there exist two irreducible representations $\mathfrak{D}^+_{\kappa,\kappa^*}$ and $\mathfrak{D}^-_{\kappa,\kappa^*}$ corresponding to the pair (κ,κ^*). The finite representations correspond to $\mathfrak{D}^-_{\kappa,\kappa^*}$ with κ,κ^* both of the form $\frac{1}{2}n(n\neq-1)$, the connexion between (p,q) and (κ,κ^*) being given by $p=|\kappa+\frac{1}{2}|-\frac{1}{2}$, $q=|\kappa^*+\frac{1}{2}|-\frac{1}{2}$. In general $\mathfrak{D}_{\kappa,\kappa^*}$ is unitary only if either

$$\kappa=-\tfrac{1}{2}+i\nu+\tfrac{1}{2}n,\quad \kappa^*=-\tfrac{1}{2}+i\nu-\tfrac{1}{2}n, \tag{1a}$$

or

$$\kappa=\kappa^*=-\tfrac{1}{2}+\nu,\quad |\nu|\leqslant\tfrac{1}{2}. \tag{1b}$$

Here ν is an arbitrary real number. However, in the special case when κ,κ^* are both of the form $\frac{1}{4}n\,(n\neq-2)$, $\mathfrak{D}^-_{\kappa,\kappa^*}$ is unitary only if $\frac{1}{2}\geqslant|\kappa+\frac{1}{2}|=|\kappa^*+\frac{1}{2}|$, while $\mathfrak{D}^+_{\kappa,\kappa^*}$ is unitary whenever $|\kappa+\frac{1}{2}|=|\kappa^*+\frac{1}{2}|$.

It is possible to construct a series in the complex variables u_1, u_2 and their complex conjugates $u_{\dot{1}}$, $u_{\dot{2}}$ such that the coefficients of this series transform according to the irreducible representation $\mathfrak{D}_{\kappa,\kappa^*}$ when a spinor transformation corresponding to a given Lorentz transformation is applied to the variables u_1, u_2 and $u_{\dot{1}}, u_{\dot{2}}$. These coefficients therefore form the components of a vector of the mathematical vector space $\mathfrak{R}_{\kappa,\kappa^*}$ which transforms according to the representation $\mathfrak{D}_{\kappa,\kappa^*}$. It is clear that these vectors of $\mathfrak{R}_{\kappa,\kappa^*}$ bear the same relation to ordinary spinors as the expansors of Dirac to ordinary tensors. They may therefore be called *expinors*. Thus to every series in $u_1, u_2, u_{\dot{1}}, u_{\dot{2}}$, which is convergent in a certain sense, there corresponds an expinor. This expinor transforms according to the representation $\mathfrak{D}_{\kappa,\kappa^*}$ if the series is homogeneous in each of the two sets of variables (u_1, u_2) and $(u_{\dot{1}}, u_{\dot{2}})$ and has the degrees 2κ and $2\kappa^*$ respectively. Such an expinor may therefore be called a homogeneous expinor. In case both 2κ and $2\kappa^*$ are integral and $\geqslant 0$, the corresponding homogeneous expinor is the same as the symmetric spinor with 2κ undotted and $2\kappa^*$ dotted indices. Homogeneous expinors are thus an extension of symmetric spinors.

---

† The term 'integer' is used throughout to mean a real integer whether positive, negative or zero. The symbol n always denotes an integer, while s denotes an integer or half odd integer $\geqslant 0$.

Further, as pointed out by Dirac (1945), it is possible to use expinors to describe the transformation properties of the wave function of a spinning particle. In a theory based on these expinors it is possible to make the charge density positive definite for particles of integral spin or the energy density positive definite for particles of half-integral spin, in contradistinction with the results of the existing theory (see Pauli 1940). This is made possible by the circumstance that infinite unitary representations of the Lorentz group exist for both integral and half-integral spins. By imposing subsidiary conditions on the wave function, it can be arranged that in the rest system the particle has a 'pure' spin s, i.e. all the non-vanishing components of the wave function in the rest system transform according to the irreducible representation $\mathfrak{D}_s$ of the rotation group.

2. THE GENERAL REPRESENTATION

The finite representations of the Lorentz group are in fact obtained from those of the four-dimensional real orthogonal group by a device which is called the 'unitarian trick' by Weyl (1939). It consists essentially in replacing in a representation three of the real parameters of the orthogonal group by pure imaginary ones. However, this device does not yield all the possible representations of the Lorentz group. This arises from the fact that the topological structure of the orthogonal group is not the same as that of the Lorentz group, the difference consisting in the aperiodic nature of the 'temporal rotations'. In consequence, unlike the orthogonal group, the Lorentz group is not compact.

We begin with the consideration of the infinitesimal transformations, which, as is well known, play a very important part in the theory of Lie groups. Such a Lorentz transformation is given by

$$x'_k = x_k + \epsilon_k{}^l x_l \quad (\epsilon_{kl} = -\epsilon_{lk}), \tag{2}$$

ϵ_{kl}'s being infinitesimal quantities. (The usual metric tensor g_{kl} is used for raising and lowering the tensor indices $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$, the other components being zero.) In any representation this transformation will be represented by

$$1 + \tfrac{1}{2}\epsilon_{kl} I^{kl}, \tag{3}$$

where I^{kl} are operators of the representation which may always be chosen to be antisymmetric in k, l. The commutation relations of I^{kl} are completely fixed by the structure of the Lorentz group (see Van der Waerden 1932) and are

$$[I^{kl}, I^{mn}] = -g^{km}I^{ln} + g^{lm}I^{kn} + g^{kn}I^{lm} - g^{ln}I^{km}, \tag{4}$$

where $[A, B] = AB - BA$. Put

$$K_1 = iI^{23}, \quad K_2 = iI^{31}, \quad K_3 = iI^{12}, \quad L_1 = iI^{10}, \quad L_2 = iI^{20}, \quad L_3 = iI^{30}. \tag{5}$$

From the definition (2) it is clear that I^{kl} transform as components of an antisymmetrical tensor, under Lorentz transformations. Therefore it follows from (5) that for spatial rotations both (K_1, K_2, K_3) and (L_1, L_2, L_3) transform as components of a three-dimensional vector. These two vectors will be denoted by $\mathbf{K}$ and $\mathbf{L}$ respectively. In terms of $\mathbf{K}$ and $\mathbf{L}$ (4) can be written as

$$[K_1, K_2] = iK_3, \tag{6a}$$

$$[K_1, L_2] = iL_3 = [L_1, K_2], \tag{6b}$$

$$[L_1, L_2] = -iK_3, \tag{6c}$$

the remaining relations being obtained by cyclic permutations of $(1, 2, 3)$.

Consider now an irreducible representation of the proper Lorentz group. The matrices I^{kl} then form an irreducible set. Denote the representation space by $\mathfrak{R}$. Reduce this space with respect to the subgroup δ_3 consisting of spatial rotations only. Now δ_3 is a compact group, and it is well known that every representation of a compact group is completely reducible into a direct sum of irreducible representations each of which is unitary and of a finite degree (see Pontrjagin 1939, p. 205). Now every irreducible representation of δ_3 is characterized by a number $k \geqslant 0$ which is integral or half-integral. Thus $\mathfrak{R}$ is decomposed into a direct sum of subspaces $\mathfrak{R}_k$ which are all irreducible with respect to δ_3. It will be assumed† that in this decomposition there occurs at most only one subspace $\mathfrak{R}_k$ for any particular value of k. By choosing a suitable coordinate system in $\mathfrak{R}_k$, K_3 can be brought to the diagonal form with its eigenvalues m running through the numbers $k, k-1, \ldots, -k+1, -k$. Further, this coordinate system in $\mathfrak{R}_k$ can always be so chosen that the matrix elements of K_1 and K_2 are given by

$$(m \mid K^+ \mid m-1) = (m-1 \mid K^- \mid m) = \sqrt{\{(k+m)(k-m+1)\}}, \tag{7}$$

where
$$K^+ = K_1 + iK_2, \quad K^- = K_1 - iK_2. \tag{8}$$

Thus K_1, K_2, K_3 are represented by Hermitian matrices and

$$\mathbf{K}^2 = k(k+1). \tag{9}$$

Now since $\mathbf{L}$ transforms as a vector under spatial rotations it follows (see, for example, Weyl 1931, p. 200) that $(k \mid \mathbf{L} \mid k') = 0$ unless $k' = k$ or $k \pm 1$. In fact, the matrix elements of **L** are completely fixed by its vector character, apart from a factor depending on k (but not on m). They are given by Weyl (1931),

---

† *Note added in proof.* It is possible to avoid this assumption. The proof is then somewhat more complicated but the final result is the same.

$$\left.\begin{aligned}(k,m\mid L^+\mid k-1,m-1) &= -p^-(k)\sqrt{\{(k+m)(k+m-1)\}},\\ (k,m-1\mid L^-\mid k-1,m) &= p^-(k)\sqrt{\{(k-m+1)(k-m)\}},\\ (k,m\mid L_3\mid k-1,m) &= p^-(k)\sqrt{\{(k+m)(k-m)\}},\end{aligned}\right\} \tag{10a}$$

$$\left.\begin{aligned}(k,m\mid L^+\mid k,m-1) &= p(k)\sqrt{\{(k+m)(k-m+1)\}},\\ (k,m-1\mid L^-\mid k,m) &= p(k)\sqrt{\{(k-m+1)(k+m)\}},\\ (k,m\mid L_3\mid k,m) &= p(k)\,m,\end{aligned}\right\} \tag{10b}$$

$$\left.\begin{aligned}(k,m\mid L^+\mid k+1,m-1) &= p^+(k)\sqrt{\{(k-m+1)(k-m+2)\}},\\ (k,m-1\mid L^-\mid k+1,m) &= -p^+(k)\sqrt{\{(k+m)(k+m+1)\}},\\ (k,m\mid L_3\mid k+1,m) &= p^+(k)\sqrt{\{(k+m+1)(k-m+1)\}},\end{aligned}\right\} \tag{10c}$$

the remaining elements vanishing. Here $p^-(k), p(k)$ and $p^+(k)$ are numerical coefficients depending on k, and $L^+ = L_1 + iL_2$, $L^- = L_1 - iL_2$. Now notice that it is possible to construct two invariants of the proper Lorentz group from I^{kl}, viz. $J^2 = -\frac{1}{2}I^{kl}I_{kl}$ and $I = -\frac{1}{2}\epsilon^{klmn}I_{kl}I_{mn}$. ($\epsilon^{klmn}$ is a tensor antisymmetric in all the indices and $\epsilon^{0123} = 1$). Since they are left unaltered by a proper Lorentz transformation, J^2 and I commute with I^{kl} and therefore in an irreducible representation they are represented by a numerical constant. Now

$$-\tfrac{1}{2}\epsilon^{klmn}I_{kl}I_{mn} = I = K_1L_1 + K_2L_2 + K_3L_3 = (\mathbf{KL}), \tag{11a}$$

$$-\tfrac{1}{2}I^{kl}I_{kl} = J^2 = \mathbf{K}^2 - \mathbf{L}^2 = k(k+1) - (L^+L^- + L_3^2 + K_3). \tag{11b}$$

Substituting from (7) and (10) in (11) one gets

$$k(k+1)\,p(k) = I, \tag{12a}$$

$$\begin{aligned}J^2 = k(k+1)(1-p^2(k)) - m(1+p^2(k)) - p^+(k)\,p^-(k+1)(k-m+1)(2k+3)\\ -p^-(k)\,p^+(k-1)(k+m)(2k-1),\end{aligned} \tag{12b}$$

it being understood that $p^-(k) \equiv 0$ if $k = 0, \frac{1}{2}$. For $k \neq 0$ there are at least two distinct values of m for which (12*b*) holds. Hence the coefficient of m must vanish for $k \neq 0$, i.e.

$$(1+p^2(k))\,k = p^+(k)\,p^-(k+1)\,k(2k+3) - p^-(k)\,p^+(k-1)\,k(2k-1), \tag{13a}$$

$$\begin{aligned}J^2 = k(k+1)\{1-p^2(k)\} - p^+(k)\,p^-(k+1)(k+1)(2k+3)\\ -p^-(k)\,p^+(k-1)\,k(2k-1).\end{aligned} \tag{13b}$$

The factor k has been introduced deliberately in (13*a*) so as to include the exceptional case $k = 0$. From (13) it follows that

$$p^+(k)\,p^-(k+1) = \{k(k+2) - k^2p^2(k) - J^2\}\frac{1}{(2k+1)(2k+3)}. \tag{14}$$

Now let j be the lowest value of k appearing in the decomposition of $\mathfrak{R}$ into $\mathfrak{R}_k$. First consider the case when $j \neq 0$. Bearing in mind that $p^-(j) = 0$ one obtains from (13) and (12a) on putting $k = j$

$$J^2 = j^2 - \frac{I^2}{j^2} - 1. \tag{15}$$

Hence from (14)

$$p^+(k)\,p^-(k+1) = \frac{(k-j+1)(k+j+1)}{(2k+1)(2k+3)}\left\{1 + \frac{I^2}{(k+1)^2 j^2}\right\} \quad (k \geqslant j). \tag{16}$$

On the other hand, if $j = 0$ then from (12a) $I = 0$ and so $p(k) = 0$, since one can always put $p(0) = 0$ without affecting the representation at all. Therefore from (14)

$$p^+(k)\,p^-(k+1) = \frac{(k+1)^2}{(2k+1)(2k+3)}\left\{1 - \frac{J^2+1}{(k+1)^2}\right\}. \tag{17}$$

(17) may be regarded as a special case of (15) and (16) for $j = 0$ provided one agrees to look upon I as $j\mu$ with $\mu = I/j$ as a definitely given number. This makes I vanish automatically when $j = 0$. Also from (12a)

$$p(k) = \frac{j\mu}{k(k+1)} \quad (k \geqslant j), \tag{18}$$

with the understanding that $p(0) \equiv 0$. By multiplying all the eigenvectors in each of the subspaces $\mathfrak{R}_k$ by a suitable normalization factor $q(k)$ it can always be arranged that

$$p^+(k) = p^-(k+1) = \sqrt{\frac{(k-j+1)(k+j+1)}{(2k+1)(2k+3)}}\sqrt{\left(1 + \frac{\mu^2}{(k+1)^2}\right)}. \tag{19}$$

For convenience the first square root is supposed to be always positive. The second one may then be chosen at will for each k. In particular, if the representation of the Lorentz group is unitary, **K** and **L** are Hermitian so that $p(k)$ is real and

$$p^+(k) = \overline{p^-(k+1)},$$

where the bar denotes complex conjugate. The normalization factors $q(k)$ are then unimodular, and the corresponding transformation of $\mathfrak{R}$ is unitary. Therefore every unitary representation of the Lorentz group can be brought into a form for which (19) holds, by means of a unitary transformation. Since unitary transformations preserve the unitary character it follows that the transformed representation is also unitary. Now (16) and (18) and (19) show that, apart from equivalence, an irreducible representation is completely fixed by the two numbers j and μ. Also as shown above a unitary representation can be transformed to the form (19) without destroying its unitary character. Hence all irreducible representations which are equivalent to a unitary representation are those for which the right-hand sides of (18) and (19) are real. *Thus for a representation to be unitary it is necessary and sufficient that* $I = \mu j$ *be real and*

$$\mu^2 > -1. \tag{20}$$

The condition (20) is actually effective only when $j = 0$.

3. THE SPINOR MATRICES $\mathfrak{U}_\alpha$, $\mathfrak{V}_\alpha$

To proceed further we must now introduce spinors. For the moment consider only the rotation group δ_3. Then, it is well known that with every vector of the three-dimensional space there can be associated a symmetrical spinor with two indices. Thus to the vector $\mathbf{K} = (K_1, K_2, K_3)$ corresponds a spinor $K_{\alpha\beta}$ defined† by

$$K_{12} = K_{21} = -K_3, \quad K_{11} = K^-, \quad K_{22} = -K^+. \tag{21}$$

As usual the antisymmetrical spinors $\epsilon^{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ $(\epsilon^{12} = \epsilon_{12} = 1)$ will be used for raising and lowering the indices according to the rule

$$a^\alpha = \epsilon^{\alpha\beta} a_\beta, \quad a_\beta = a^\alpha \epsilon_{\alpha\beta}. \tag{22}$$

Now it is well known that $K_{\alpha\beta}(k)$ (which is the matrix representing $K_{\alpha\beta}$ in $\mathfrak{R}_k$) can be expressed in terms of the spinor matrices $u_\alpha(k), v_\beta(k)$ introduced by Dirac (1936) and Fierz (1939). $u_1(k), u_2(k)$ have $2k+1$ rows and $2k$ columns and $v_1(k), v_2(k)$ have $2k$ rows and $2k+1$ columns. Their properties may briefly be summarized as follows:

$$u_\alpha(k)\, v^\alpha(k) = 2k, \quad v^\alpha(k)\, u_\alpha(k) = 2k+1, \tag{23a}$$

$$u_\alpha(k+\tfrac{1}{2})\, u^\alpha(k) = v^\alpha(k)\, v_\alpha(k+\tfrac{1}{2}) = 0, \tag{23b}$$

$$v^\alpha(k) = \overline{u_\alpha(k)}, \quad u_\alpha(k) = \overline{v^\alpha(k)}, \tag{23c}$$

$$v_\alpha(k+\tfrac{1}{2})\, u_\beta(k+\tfrac{1}{2}) - u_\beta(k)\, v_\alpha(k) = \epsilon_{\beta\alpha}, \tag{23d}$$

$$K_{\alpha\beta}(k) = u_\alpha(k)\, v_\beta(k) - k\epsilon_{\alpha\beta} = v_\alpha(k+\tfrac{1}{2})\, u_\beta(k+\tfrac{1}{2}) + (k+1)\,\epsilon_{\alpha\beta}. \tag{23e}$$

The bar denotes Hermitian conjugate. It is easy to show that on account of the Hermitian character of $\mathbf{K}$, $u_\alpha(k)$ and $v_\beta(k)$ can always be chosen to satisfy (23*c*). Notice that (23*c*) is a proper spinor equation, since for spatial rotations a lower spinor index transforms the same way as the complex conjugate of an upper spinor index.

Now define two matrices U^α and V^α as follows:

$$(k \mid U^\alpha \mid k-\tfrac{1}{2}) = u^\alpha(k), \quad (k-\tfrac{1}{2} \mid V^\alpha \mid k) = v^\alpha(k), \tag{24}$$

all the other elements being zero. Then clearly (23) can be written as

$$U_\alpha V^\alpha = 2k, \quad V^\alpha U_\alpha = 2(k+1), \tag{25a}$$

$$U_\alpha U^\alpha = V^\alpha V_\alpha = 0, \tag{25b}$$

$$V^\alpha = \overline{U_\alpha}, \quad U_\alpha = \overline{V^\alpha}, \tag{25c}$$

$$V_\alpha U_\beta - U_\beta V_\alpha = \epsilon_{\beta\alpha}, \tag{25d}$$

$$K_{\alpha\beta} = U_\alpha V_\beta - k\epsilon_{\alpha\beta} = V_\alpha U_\beta + (k+1)\,\epsilon_{\alpha\beta}. \tag{25e}$$

Here k is the diagonal matrix whose component inside $\mathfrak{R}_{k'}$ is simply k' times the unit matrix. Notice that from (24)

$$V^\alpha k = (k+\tfrac{1}{2})\, V^\alpha, \tag{26a}$$

$$U^\alpha k = (k-\tfrac{1}{2})\, U^\alpha. \tag{26b}$$

---

† Spinor indices are always denoted by Greek letters and tensor indices by Latin letters.

Now construct similarly a spinor $L_{\alpha\beta}$ corresponding to the vector $\mathbf{L}$,

$$L_{12} = L_{21} = -L_3, \quad L_{11} = L^-, \quad L_{22} = -L^+. \tag{27}$$

Then (6) can be written as

$$[K_{\alpha\beta}, K_{\lambda\mu}] = \tfrac{1}{2} \underset{(\alpha\beta)}{S} \; \underset{(\lambda\mu)}{S} \, \epsilon_{\lambda\alpha} K_{\mu\beta}, \tag{28a}$$

$$[K_{\alpha\beta}, L_{\lambda\mu}] = [L_{\alpha\beta}, K_{\lambda\mu}] = \tfrac{1}{2} \underset{(\alpha\beta)}{S} \; \underset{(\lambda\mu)}{S} \, \epsilon_{\lambda\alpha} L_{\mu\beta}, \tag{28b}$$

$$[L_{\alpha\beta}, L_{\lambda\mu}] = -\tfrac{1}{2} \underset{(\alpha\beta)}{S} \; \underset{(\lambda\mu)}{S} \, \epsilon_{\lambda\alpha} K_{\mu\beta}, \tag{28c}$$

where $\underset{(\alpha\beta)}{S}$ denotes that terms arising from the interchange of α and β are to be added. Also

$$-\tfrac{1}{2} K^{\alpha\beta} K_{\alpha\beta} = \mathbf{K}^2, \quad -\tfrac{1}{2} L^{\alpha\beta} L_{\alpha\beta} = \mathbf{L}^2, \tag{29a}$$

$$-\tfrac{1}{2} K^{\alpha\beta} L_{\alpha\beta} = (\mathbf{KL}) = I. \tag{29b}$$

Now it follows from (10) and (19) that

$$L_{\alpha\beta} = \frac{I}{k(k+1)} K_{\alpha\beta} + i \sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}} \sqrt{\left(1+\frac{\mu^2}{k^2}\right)} U_\alpha U_\beta$$
$$+ V_\alpha V_\beta i \sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}} \sqrt{\left(1+\frac{\mu^2}{k^2}\right)}. \tag{30}$$

As before, the matrix square root $\sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}}$ is always taken to be positive while $\sqrt{\left(1+\frac{\mu^2}{k^2}\right)}$ is defined arbitrarily. To verify (30) observe that (28*b*) follows from the spinor character of $L_{\alpha\beta}$ while (28*c*) is easily obtained on using (30) and (24). Similarly (29*b*) and (15) are also easily verified. This fixes the j and μ value of the representation uniquely, and therefore as shown in § 2, (29) must lead to the same representation, apart from equivalence, as (18) and (19), i.e. if $u_\alpha(k)$ and $v_\alpha(k)$ are suitably chosen the two representations must coincide.

Now put

$$I_{\alpha\beta} = \tfrac{1}{2}(K_{\alpha\beta} - iL_{\alpha\beta}), \tag{31a}$$

and notice that

$$[I_{\alpha\beta}, I_{\lambda\mu}] = \tfrac{1}{2} \underset{(\alpha\beta)}{S} \; \underset{(\lambda\mu)}{S} \, \epsilon_{\lambda\alpha} I_{\mu\beta}, \tag{31b}$$

$$-\tfrac{1}{2} I^{\alpha\beta} I_{\alpha\beta} = \frac{J^2 - 2iI}{4} = \tfrac{1}{4}(j - i\mu)^2 - \tfrac{1}{4}. \tag{31c}$$

Consider the product space $\mathfrak{R} \times \mathfrak{R}_{\frac{1}{2}}$, where $\mathfrak{R}_{\frac{1}{2}}$ is a two-dimensional space transforming according to the fundamental spinor representation $\mathfrak{D}_{\frac{1}{2},0}$ of the Lorentz group. The components of a vector in this space may be written as ψ_α ($\alpha = 1, 2$), where ψ_1 and ψ_2 are one-column matrices enumerating the components in $\mathfrak{R}$. With respect to the index α, ψ_α transforms as a spinor. Now define

$$\psi^* = U^\alpha \frac{1}{\sqrt{(2k+1)}} \psi_\alpha, \tag{32a}$$

$$\psi = V^\alpha \frac{1}{\sqrt{(2k+1)}} \psi_\alpha. \tag{32b}$$

Then from (25) $$\psi_\alpha = \frac{1}{\sqrt{(2k+1)}} U_\alpha \psi - \frac{1}{\sqrt{(2k+1)}} V_\alpha \psi^*. \tag{32c}$$

Now put $$I_\lambda{}^\mu \psi_\mu = \psi'_\lambda. \tag{33}$$

From (31*a*) and (30) $$I_{\alpha\beta} = a(k) K_{\alpha\beta} + b(k) U_\alpha U_\beta + V_\alpha V_\beta b(k), \tag{34a}$$

where $$a(k) = \frac{1}{2}\left(1 - \frac{iI}{k(k+1)}\right), \quad b(k) = \frac{1}{2}\sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}}\sqrt{\left(1+\frac{\mu^2}{k^2}\right)}. \tag{34b}$$

On using (25), (26), (32) and (34) one finds that

$$\psi' \equiv V^\lambda \cdot \frac{1}{\sqrt{(2k+1)}} \psi'_\lambda = a(k+\tfrac{1}{2})(k+\tfrac{3}{2})\psi + b(k+\tfrac{1}{2})\sqrt{\{2k(2k+2)\}}\,\psi^*, \tag{35a}$$

and similarly

$$\psi^{*\prime} \equiv U^\lambda \frac{1}{\sqrt{(2k+1)}} \psi'_\lambda = -a(k-\tfrac{1}{2})(k-\tfrac{1}{2})\psi^* - b(k+\tfrac{1}{2})\sqrt{\{2k(2k+2)\}}\,\psi. \tag{35b}$$

(35) may be looked upon as a linear transformation in the two 'variables' ψ and ψ^*. The eigenvalues of this transformation are easily obtained on using (34*b*). They are $\frac{1}{2}(1 \pm \gamma)$, where γ is a complex number defined, except for an ambiguity in sign, by the equation

$$\gamma^2 = (j - i\mu)^2. \tag{36}$$

Write $\gamma_{k+\frac{1}{2}} = k + \frac{1}{2} - \frac{iI}{k+\frac{1}{2}}$, and put

$$\sqrt{2}\,\psi_{\gamma-1} = \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi + \sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi^*, \tag{37a}$$

$$\sqrt{2}\,\psi_{\gamma+1} = \sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi + \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi^*, \tag{37b}$$

the square roots being so chosen that

$$\sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)} = \sqrt{\{(k+\tfrac{1}{2}-j)(k+\tfrac{1}{2}+j)\}}\sqrt{\left(1+\frac{\mu^2}{(k+\frac{1}{2})^2}\right)}. \tag{37c}$$

From (35) and (37) $$\sqrt{2}\,\psi'_{\gamma-1} \equiv \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi' + \sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi^{*\prime}$$
$$= \tfrac{1}{2}(1+\gamma)\sqrt{2}\,\psi_{\gamma-1}. \tag{38a}$$

Similarly $$\sqrt{2}\,\psi'_{\gamma+1} \equiv \sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi' + \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi^{*\prime}$$
$$= \tfrac{1}{2}(1-\gamma)\sqrt{2}\,\psi_{\gamma+1}. \tag{38b}$$

Thus the transformation (37) brings (35) to the diagonal form. Also it follows from (37) that

$$\sqrt{2}\,\gamma\psi = \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi_{\gamma-1} - \sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi_{\gamma+1}, \tag{39a}$$

$$\sqrt{2}\,\gamma\psi^* = -\sqrt{(\gamma_{k+\frac{1}{2}} - \gamma)}\,\psi_{\gamma-1} + \sqrt{(\gamma_{k+\frac{1}{2}} + \gamma)}\,\psi_{\gamma+1}. \tag{39b}$$

Now corresponding to (36) define γ^* by the equation

$$\gamma^{*2} = (j + i\mu)^2, \tag{40}$$

and write $\gamma = 2\kappa+1, \gamma^* = 2\kappa^*+1$. Then

$$iI = \frac{\gamma^{*2}-\gamma^2}{4} = \kappa^*(\kappa^*+1)-\kappa(\kappa+1) = (\kappa^*-\kappa)(\kappa^*+\kappa+1), \tag{41}$$

$$\gamma_{k+\frac{1}{2}} = k+\tfrac{1}{2}-\frac{iI}{k+\frac{1}{2}} = \frac{(k+\frac{1}{2})^2-(\kappa^*+\frac{1}{2})^2+(\kappa+\frac{1}{2})^2}{(k+\frac{1}{2})}.$$

Hence
$$\gamma_{k+\frac{1}{2}}+\gamma = \frac{(k+\kappa+1)^2-(\kappa^*+\frac{1}{2})^2}{(k+\frac{1}{2})} = \frac{(k+\kappa+\kappa^*+\frac{3}{2})(k+\kappa-\kappa^*+\frac{1}{2})}{(k+\frac{1}{2})}, \tag{42a}$$

$$\gamma_{k+\frac{1}{2}}-\gamma = \frac{(k-\kappa)^2-(\kappa^*+\frac{1}{2})^2}{k+\frac{1}{2}} = \frac{(k-\frac{1}{2}-\kappa-\kappa^*)(k+\frac{1}{2}-\kappa+\kappa^*)}{(k+\frac{1}{2})}. \tag{42b}$$

Put $\psi_{\gamma-1} \equiv \psi_{\kappa-\frac{1}{2}}, \psi_{\gamma+1} \equiv \psi_{\kappa+\frac{1}{2}}$. Then it follows from (37) and (32) that

$$\psi_{\kappa-\frac{1}{2}} = \mathfrak{B}^\alpha(\kappa,\kappa^*)\,\psi_\alpha, \tag{43a}$$

$$\psi_{k+\frac{1}{2}} = \mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_\alpha, \tag{43b}$$

where $\mathfrak{B}^\alpha$ and $\mathfrak{U}^\alpha$ are defined by

$$\mathfrak{U}^\alpha(\kappa,\kappa^*) = \tfrac{1}{2}V^\alpha \sqrt{\left\{\frac{(k+1+\kappa+\kappa^*)(k+\kappa-\kappa^*)}{k(k+\frac{1}{2})}\right\}} + \frac{1}{2}\sqrt{\left\{\frac{(k+\frac{1}{2}-\kappa+\kappa^*)(k-\frac{1}{2}-\kappa-\kappa^*)}{k(k+\frac{1}{2})}\right\}}\,U^\alpha, \tag{44a}$$

$$\mathfrak{B}^\alpha(\kappa,\kappa^*) = \tfrac{1}{2}V^\alpha \sqrt{\left\{\frac{(k+\frac{1}{2}-\kappa+\kappa^*)(k-\frac{1}{2}-\kappa-\kappa^*)}{k(k+\frac{1}{2})}\right\}} + \frac{1}{2}\sqrt{\left\{\frac{(k+1+\kappa+\kappa^*)(k+\kappa-\kappa^*)}{k(k+\frac{1}{2})}\right\}}\,U^\alpha. \tag{44b}$$

Further from (32*c*), (39) and (44) it is found that

$$(2\kappa+1)\,\psi_\alpha = \mathfrak{U}_\alpha(\kappa,\kappa^*)\,\psi_{k-\frac{1}{2}} - \mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_{\kappa+\frac{1}{2}}. \tag{45}$$

When $\gamma = 2\kappa+1 \neq 0$ the two eigenvalues $\frac{1}{2}(1\pm\gamma)$ of (35) are distinct and hence the two 'eigenvectors' $\psi_{\kappa-\frac{1}{2}}$ and $\psi_{\kappa+\frac{1}{2}}$ are linearly independent. Now from (43) and (45)

$$(2\kappa+1)\,\psi_\alpha = \mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{B}^\beta(\kappa,\kappa^*)\,\psi_\beta - \mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}^\beta(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_\beta, \tag{46a}$$

$$(2\kappa+1)\,\psi_{k-\frac{1}{2}} = \mathfrak{B}^\alpha(\kappa,\kappa^*)\,\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\psi_{k-\frac{1}{2}} - \mathfrak{B}^\alpha(\kappa,\kappa^*)\,\mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_{k+\frac{1}{2}}, \tag{46b}$$

$$(2\kappa+1)\,\psi_{\kappa+\frac{1}{2}} = \mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\psi_{\kappa-\frac{1}{2}} - \mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_{\kappa+\frac{1}{2}}. \tag{46c}$$

Also from (33), (38) and (43)

$$(\kappa+1)\,\psi_{\kappa-\frac{1}{2}} = \mathfrak{B}^\alpha(\kappa,\kappa^*)\,I_\alpha{}^\beta\psi_\beta, \tag{47a}$$

$$-\kappa\psi_{\kappa+\frac{1}{2}} = \mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,I_\alpha{}^\beta\psi_\beta, \tag{47b}$$

so that from (45)

$$\begin{aligned}(2\kappa+1)\,I_\alpha{}^\beta\psi_\beta &= (\kappa+1)\,\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\psi_{\kappa-\frac{1}{2}} + \kappa\mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_{\kappa+\frac{1}{2}} \\ &= (\kappa+1)\,\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{B}^\beta(\kappa,\kappa^*)\,\psi_\beta + \kappa\mathfrak{B}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}^\beta(\kappa+\tfrac{1}{2},\kappa^*)\,\psi_\beta.\end{aligned} \tag{48}$$

Now ψ_α is completely arbitrary. Therefore from (46*a*) and (48)

$$(2\kappa+1)\,\epsilon_\alpha{}^\beta = \mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{V}^\beta(\kappa,\kappa^*) - \mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}^\beta(\kappa+\tfrac{1}{2},\kappa^*), \tag{49a}$$

$$(2\kappa+1)\,I_\alpha{}^\beta = (\kappa+1)\,\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{V}^\beta(\kappa,\kappa^*) + \kappa\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}^\beta(\kappa+\tfrac{1}{2},\kappa^*). \tag{49b}$$

Now when $\gamma = 2\kappa+1 \neq 0$ the transformation (37) is non-singular since it has an inverse, namely, (39). Hence the correspondence between ψ_α and the pair $(\psi_{\kappa-\frac{1}{2}}, \psi_{\kappa+\frac{1}{2}})$ is a 1-1 correspondence. Thus every arbitrarily chosen pair $(\psi_{\kappa-\frac{1}{2}}, \psi_{\kappa+\frac{1}{2}})$ corresponds to a unique ψ_α, and therefore (46) and (47) hold for arbitrary $\psi_{\kappa-\frac{1}{2}}$ and $\psi_{\kappa+\frac{1}{2}}$ if $2\kappa+1 \neq 0$. This implies that

$$\mathfrak{V}^\alpha(\kappa,\kappa^*)\,\mathfrak{U}_\alpha(\kappa,\kappa^*) = 2\kappa+1, \quad \mathfrak{V}^\alpha(\kappa,\kappa^*)\,\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*) = 0,$$

$$\mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}_\alpha(\kappa,\kappa^*) = 0, \quad \mathfrak{U}^\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*) = -2\kappa-1.$$

After a little algebraic manipulation all these relations can be expressed as follows:

$$\mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{V}^\alpha(\kappa,\kappa^*) = 2\kappa, \quad \mathfrak{V}^\alpha(\kappa,\kappa^*)\,\mathfrak{U}_\alpha(\kappa,\kappa^*) = 2\kappa+1, \tag{50a}$$

$$\mathfrak{U}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}^\alpha(\kappa,\kappa^*) = \mathfrak{V}^\alpha(\kappa,\kappa^*)\,\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*) = 0, \tag{50b}$$

$$\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}_\beta(\kappa+\tfrac{1}{2},\kappa^*) - \mathfrak{U}_\beta(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*) = \epsilon_{\beta\alpha}, \tag{50c}$$

$$I_{\alpha\beta} = \mathfrak{U}_\alpha(\kappa,\kappa^*)\,\mathfrak{V}_\beta(\kappa,\kappa^*) - \kappa\epsilon_{\alpha\beta} = \mathfrak{V}_\alpha(\kappa+\tfrac{1}{2},\kappa^*)\,\mathfrak{U}_\beta(\kappa+\tfrac{1}{2},\kappa^*) + (\kappa+1)\,\epsilon_{\alpha\beta}. \tag{50d}$$

For the exceptional case $\kappa = -\frac{1}{2}$ (50) can be verified directly from (44) and (34). Thus the analogy between (50) and (23) is complete. The following relations are also worth noting:

$$\mathfrak{U}^\alpha(\kappa,\kappa^*) = \mathfrak{U}^\alpha(\kappa,-\kappa^*-1), \quad \mathfrak{V}^\alpha(\kappa,\kappa^*) = \mathfrak{V}^\alpha(\kappa,-\kappa^*-1); \tag{51a}$$

$$\mathfrak{V}^\alpha(-\kappa-\tfrac{1}{2},\kappa^*) = \mathfrak{U}^\alpha(\kappa,\kappa^*), \quad \mathfrak{U}^\alpha(-\kappa-\tfrac{1}{2},\kappa^*) = \mathfrak{V}^\alpha(\kappa,\kappa^*). \tag{51b}$$

Now put

$$U_\alpha = \overline{V^\alpha} = iU^{\dot\alpha}, \quad V^\alpha = \overline{U_\alpha} = iV_{\dot\alpha}. \tag{52a}$$

The equations (25*a*) to (25*d*) then hold unchanged with dotted indices provided these are raised and lowered by means of the spinors $\epsilon^{\dot\alpha\dot\beta}$, $\epsilon_{\dot\alpha\dot\beta}(\epsilon^{\dot1\dot2} = \epsilon_{\dot1\dot2} = 1)$.

Also

$$K_{\alpha\beta} = U_\alpha V_\beta - k\epsilon_{\alpha\beta} = U^{\dot\alpha}V^{\dot\beta} - k\epsilon^{\dot\alpha\dot\beta} \equiv K^{\dot\alpha\dot\beta} \tag{52b}$$

and

$$\begin{aligned} L_{\alpha\beta} = \frac{I}{k(k+1)}K^{\dot\alpha\dot\beta} - i\sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}}\sqrt{\left(1+\frac{\mu^2}{k^2}\right)}\,U^{\dot\alpha}U^{\dot\beta} \\ - V^{\dot\alpha}V^{\dot\beta}i\sqrt{\frac{(k-j)(k+j)}{(2k-1)(2k+1)}}\sqrt{\left(1+\frac{\mu^2}{k^2}\right)} \\ \equiv L^{\dot\alpha\dot\beta}. \end{aligned} \tag{52c}$$

Put

$$I^{\dot\alpha\dot\beta} = \tfrac{1}{2}(K^{\dot\alpha\dot\beta} + iL^{\dot\alpha\dot\beta}). \tag{53a}$$

Then

$$[I_{\alpha\beta}, I_{\dot\lambda\dot\mu}] = 0, \tag{53b}$$

$$[I_{\dot\alpha\dot\beta}, I_{\dot\lambda\dot\mu}] = \tfrac{1}{2}\underset{(\dot\alpha\dot\beta)}{S}\,\underset{(\dot\lambda\dot\mu)}{S}\,\epsilon_{\dot\lambda\dot\alpha}I_{\dot\mu\dot\beta}, \tag{53c}$$

$$\begin{aligned} -\tfrac{1}{2}I^{\dot\alpha\dot\beta}I_{\dot\alpha\dot\beta} = \tfrac{1}{4}(J^2+2iI) = \tfrac{1}{4}(j+i\mu)^2 - \tfrac{1}{4} = \frac{\gamma^{*2}-1}{4} \\ = \kappa^*(\kappa^*+1). \end{aligned} \tag{53d}$$

Now consider the product space $\mathfrak{R} \times \mathfrak{R}_{\frac{1}{2}}$ where $\mathfrak{R}_{\frac{1}{2}}$ is a two-dimensional space whose vectors transform as dotted spinors under transformations of the Lorentz group. As before, the components of a vector of the product space may be written as $\psi_{\dot{\mu}}$. Put

$$\dot{\psi}^* = U^{\dot{\alpha}} \frac{1}{\sqrt{(2k+1)}} \psi_{\dot{\alpha}}, \tag{54a}$$

$$\dot{\psi} = V^{\dot{\alpha}} \frac{1}{\sqrt{(2k+1)}} \psi_{\dot{\alpha}}. \tag{54b}$$

Then

$$\psi_{\dot{\alpha}} = \frac{1}{\sqrt{(2k+1)}} U_{\dot{\alpha}} \dot{\psi} - \frac{1}{\sqrt{(2k+1)}} V_{\dot{\alpha}} \dot{\psi}^*. \tag{54c}$$

As before, consider $\psi'_{\dot{\lambda}} = I_{\dot{\lambda}}^{\dot{\mu}} \psi_{\dot{\mu}}$ and construct $\dot{\psi}'$, $\dot{\psi}^{*\prime}$. Write $\gamma^*_{k+\frac{1}{2}} = k + \frac{1}{2} + \frac{iI}{k+\frac{1}{2}}$ and put

$$\sqrt{2}\, \psi_{\gamma^*-1} = \sqrt{(\gamma^*_{k+\frac{1}{2}} + \gamma^*)}\, \dot{\psi} + \sqrt{(\gamma^*_{k+\frac{1}{2}} - \gamma^*)}\, \dot{\psi}^*, \tag{55a}$$

$$\sqrt{2}\, \psi_{\gamma^*+1} = \sqrt{(\gamma^*_{k+\frac{1}{2}} - \gamma^*)}\, \dot{\psi} + \sqrt{(\gamma^*_{k+\frac{1}{2}} + \gamma^*)}\, \dot{\psi}^*, \tag{55b}$$

the square roots being so chosen that

$$\sqrt{(\gamma^*_{k+\frac{1}{2}} + \gamma^*)}\, \sqrt{(\gamma^*_{k+\frac{1}{2}} - \gamma^*)} = \sqrt{\{(k+\tfrac{1}{2}-j)\,(k+\tfrac{1}{2}+j)\}} \Big/ \sqrt{\left(1 + \frac{\mu^2}{k^2}\right)}.$$

From (55) one finds that

$$\sqrt{2}\, \psi'_{\gamma^*-1} = \tfrac{1}{2}(1+\gamma^*)\sqrt{2}\, \psi_{\gamma^*-1}, \quad \sqrt{2}\, \psi'_{\gamma^*+1} = \tfrac{1}{2}(1-\gamma^*)\sqrt{2}\, \psi_{\gamma^*+1}.$$

Put $\psi_{\gamma^*-1} = \psi_{\kappa^*-\frac{1}{2}}$, $\psi_{\gamma^*+1} = \psi_{\kappa^*+\frac{1}{2}}$, and write

$$\mathfrak{B}^{\dot{\alpha}}(\kappa, \kappa^*) = \tfrac{1}{2} V^{\dot{\alpha}} \sqrt{\left(\frac{(k+1+\kappa+\kappa^*)\,(k+\kappa^*-\kappa)}{k(k+\frac{1}{2})}\right)} + \frac{1}{2} \sqrt{\left(\frac{(k+\frac{1}{2}-\kappa^*+\kappa)\,(k-\frac{1}{2}-\kappa^*-\kappa)}{k(k+\frac{1}{2})}\right)}\, U^{\dot{\alpha}}, \tag{56a}$$

$$\mathfrak{U}^{\dot{\alpha}}(\kappa, \kappa^*) = \tfrac{1}{2} V^{\dot{\alpha}} \sqrt{\left(\frac{(k+\frac{1}{2}-\kappa^*+\kappa)\,(k-\frac{1}{2}-\kappa^*-\kappa)}{k(k+\frac{1}{2})}\right)} + \frac{1}{2} \sqrt{\left(\frac{(k+1+\kappa+\kappa^*)\,(k+\kappa^*-\kappa)}{k(k+\frac{1}{2})}\right)}\, U^{\dot{\alpha}}, \tag{56b}$$

then

$$\psi_{\kappa^*-\frac{1}{2}} = \mathfrak{B}^{\dot{\alpha}}(\kappa, \kappa^*)\, \psi_{\dot{\alpha}}, \tag{57a}$$

$$\psi_{\kappa^*+\frac{1}{2}} = \mathfrak{U}^{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2})\, \psi_{\dot{\alpha}}, \tag{57b}$$

$$(2\kappa^*+1)\, \psi_{\dot{\alpha}} = \mathfrak{U}_{\dot{\alpha}}(\kappa, \kappa^*)\, \psi_{\kappa^*-\frac{1}{2}} - \mathfrak{B}_{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2})\, \psi_{\kappa^*+\frac{1}{2}}. \tag{57c}$$

Corresponding to (50) one gets

$$\mathfrak{U}_{\dot{\alpha}}(\kappa, \kappa^*)\, \mathfrak{B}^{\dot{\alpha}}(\kappa, \kappa^*) = 2\kappa^*, \quad \mathfrak{B}^{\dot{\alpha}}(\kappa, \kappa^*)\, \mathfrak{U}_{\dot{\alpha}}(\kappa, \kappa^*) = 2\kappa^*+1, \tag{58a}$$

$$\mathfrak{U}_{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2})\, \mathfrak{U}^{\dot{\alpha}}(\kappa, \kappa^*) = \mathfrak{B}^{\dot{\alpha}}(\kappa, \kappa^*)\, \mathfrak{B}_{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2}) = 0, \tag{58b}$$

$$\mathfrak{B}_{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2})\, \mathfrak{U}_{\dot{\beta}}(\kappa, \kappa^*+\tfrac{1}{2}) - \mathfrak{U}_{\dot{\beta}}(\kappa, \kappa^*)\, \mathfrak{B}_{\dot{\alpha}}(\kappa, \kappa^*) = \epsilon_{\dot{\beta}\dot{\alpha}}, \tag{58c}$$

$$I_{\dot{\alpha}\dot{\beta}} = \mathfrak{U}_{\dot{\alpha}}(\kappa, \kappa^*)\, \mathfrak{B}_{\dot{\beta}}(\kappa, \kappa^*) - \kappa^* \epsilon_{\alpha\beta} = \mathfrak{B}_{\dot{\alpha}}(\kappa, \kappa^*+\tfrac{1}{2})\, \mathfrak{U}_{\dot{\beta}}(\kappa, \kappa^*+\tfrac{1}{2}) + (\kappa^*+1)\, \epsilon_{\alpha\beta}, \tag{58d}$$

and
$$\mathfrak{U}^{\dot\alpha}(\kappa,\kappa^*) = \mathfrak{U}^{\dot\alpha}(-\kappa-1,\kappa^*), \quad \mathfrak{B}^{\dot\alpha}(\kappa,\kappa^*) = \mathfrak{B}^{\dot\alpha}(-\kappa-1,\kappa^*), \tag{59a}$$
$$\mathfrak{U}^{\dot\alpha}(\kappa, -\kappa^*-\tfrac{1}{2}) = \mathfrak{B}^{\dot\alpha}(\kappa,\kappa^*), \quad \mathfrak{B}^{\dot\alpha}(\kappa, -\kappa^*-\tfrac{1}{2}) = \mathfrak{U}^{\dot\alpha}(\kappa,\kappa^*). \tag{59b}$$

Now let the numbers γ and γ^* be given for a particular representation. Then from (36) and (40)
$$j = \frac{\epsilon\gamma+\epsilon^*\gamma^*}{2}, \quad i\mu = \frac{\epsilon^*\gamma^*-\epsilon\gamma}{2}, \quad iI = \frac{\gamma^{*2}-\gamma^2}{4}, \tag{60}$$
where $\epsilon = \pm 1, \epsilon^* = \pm 1$, the two ambiguities being independent. Since $2j$ is integral and $\geqslant 0$ it follows that at least one of the two quantities $\gamma \pm \gamma^*$ must be an integer. In case only one of these quantities is integral the value of j is completely determined from those of γ and γ^*. This determines ϵ and ϵ^* and therefore μ except in the case when $j = 0$. But even in this case I and μ^2 are completely determined. Since only j, I and μ^2 enter in the representation it follows that the representation is completely determined by the pair (γ,γ^*). However, when both $\gamma+\gamma^*$ and $\gamma-\gamma^*$ are integral and distinct there are two possible values of j, say j^+ and j^-, such that $j^+ > j^-$. Therefore in this case there are two distinct representations corresponding to the pair (γ,γ^*). They may be denoted by $\mathfrak{D}^+_{\kappa,\kappa^*}$ and $\mathfrak{D}^-_{\kappa,\kappa^*}$ respectively, corresponding to the j value j^+ and j^-. Thus a given pair (κ,κ^*) completely determines the irreducible representation unless κ,κ^* are both of the form $\frac{1}{2}n$ $(n \neq -2)$. It is easy to see that $\mathfrak{D}^+_{\kappa,\kappa^*}$ is always an infinite representation. $\mathfrak{D}^-_{\kappa,\kappa^*}$ is finite only if j^- and j^+ differ by an integer which is the case when both κ and κ^* are of the form $\frac{1}{2}n$ $(n \neq -1)$. Since reversing the sign of γ or γ^* does not alter the representation the four pairs (κ,κ^*), $(-\kappa-1,\kappa^*)$, $(\kappa,-\kappa^*-1)$, $(-\kappa-1,-\kappa^*-1)$ determine the same representation. From (60) it follows that at least one of the two quantities $\kappa-\kappa^*$ and $\kappa+\kappa^*+1$ is of the form $\frac{1}{2}n$. It may therefore be fixed for convenience that κ,κ^* are to be so chosen that $\kappa-\kappa^*$ is an integer or half-odd integer. This is the only condition to be imposed on the choice of κ,κ^*.

It is useful to examine the procedure which led to the matrices $\mathfrak{U}_\alpha$, $\mathfrak{B}_\alpha$ and $\mathfrak{U}_{\dot\alpha}$, $\mathfrak{B}_{\dot\alpha}$ a little more closely. From (15), (36) and (40)
$$\gamma^2 = J^2+1-2iI, \quad \gamma^{*2} = J^2+i+2iI. \tag{61}$$
(61) may be used to define γ and γ^* as operators in terms of $\mathbf{K}$ and $\mathbf{L}$ even for those reducible representations in which J^2 and I are no longer represented by multiples of the unit matrix. In particular in the product space $\mathfrak{R}\times\mathfrak{R}_{\frac{1}{2}}$ the infinitesimal transformations $\mathfrak{K}$ and $\mathfrak{L}$ corresponding to $\mathbf{K}$ and $\mathbf{L}$ are given by
$$\mathfrak{K} = \mathbf{K}+\tfrac{1}{2}\boldsymbol{\sigma}, \quad \mathfrak{L} = \mathbf{L}+\tfrac{1}{2}\boldsymbol{\sigma},$$
where $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\sigma_3)$ is the vector constructed from the three Pauli matrices $(\sigma_1\sigma_2 = i\sigma_3)$. Denote by $\mathscr{I}$ and $\mathscr{J}^2$ the operators in the product space corresponding to I and J^2 so that
$$\mathscr{I} = (\mathfrak{K}\mathfrak{L}) = (\mathbf{K}+\tfrac{1}{2}\boldsymbol{\sigma})(\mathbf{L}+\tfrac{1}{2}i\boldsymbol{\sigma}) = I+\tfrac{3}{4}i+iA, \tag{62a}$$
$$\mathscr{J}^2 = \mathfrak{K}^2-\mathfrak{L}^2 = (\mathbf{K}+\tfrac{1}{2}\boldsymbol{\sigma})^2-(\mathbf{L}+\tfrac{1}{2}i\boldsymbol{\sigma})^2 = J^2+\tfrac{3}{2}+2A, \tag{62b}$$
where
$$A = \tfrac{1}{2}(\mathbf{K}\,.\boldsymbol{\sigma})-\tfrac{1}{2}i(\mathbf{L}\,.\,\boldsymbol{\sigma}). \tag{62c}$$

Now notice that A is an invariant since

$$A = -\tfrac{1}{2} I^{kl} I_{kl}(\tfrac{1}{2}),$$

where I^{kl} and $I^{kl}(\frac{1}{2})$ are the infinitesimal transformations of the spaces $\mathfrak{R}$ and $\mathfrak{R}_{\frac{1}{2}}$ respectively. Hence A commutes with $\boldsymbol{\mathfrak{K}}$ and $\boldsymbol{\mathfrak{L}}$. Therefore in an irreducible subspace of $\mathfrak{R} \times \mathfrak{R}_{\frac{1}{2}}$ it must be represented by a multiple of the unit matrix. From (62c) one finds that

$$A^2 = \tfrac{1}{4}(J^2 - 2iI) - A,$$

i.e.

$$(A + \tfrac{1}{2})^2 = \tfrac{1}{4}(J^2 - 2iI + 1) = \tfrac{1}{4}\gamma^2.$$

So that the eigenvalues of A are $-\frac{1}{2}(1 \pm \gamma)$. If Γ and Γ^* denote the operators of the product space corresponding to γ and γ^* respectively, then

$$\Gamma^{*2} = \mathscr{J}^2 + 2i\mathscr{J} + 1 = J^2 + 1 + 2iI = \gamma^{*2}, \tag{63a}$$

$$\Gamma^2 = \mathscr{J}^2 - 2i\mathscr{J} + 1 = J^2 + 1 - 2iI + 4A + 3 = \gamma^2 + 4A + 3. \tag{63b}$$

Hence for the two irreducible subspaces corresponding to the two eigenvalues of A one gets

$$\Gamma^2 = (\gamma \mp 1)^2.$$

Hence one may choose $\Gamma^* = \gamma^*$, $\Gamma = \gamma \mp 1$, the choice of sign being dependent on the corresponding choice in the eigenvalues $-\frac{1}{2}(1 \pm \gamma)$ of A. $-A$ is precisely the transformation $\psi_\mu \rightarrow \psi'_\mu = I_\mu{}^\lambda \psi_\lambda$ (the negative sign coming from the fact that the spinor component $K_{12} = -K_3$). Hence it is clear that $\psi_{\kappa\pm\frac{1}{2}}$ and $\psi_{\kappa^*\pm\frac{1}{2}}$ transform according to the representations $(\kappa \pm \frac{1}{2}, \kappa^*)$ and $(\kappa, \kappa^* \pm \frac{1}{2})$ respectively. In view of this it is more logical to write $\psi(\kappa \pm \frac{1}{2}, \kappa^*)$ and $\psi(\kappa, \kappa^* \pm \frac{1}{2})$ respectively. The j values of these representations are given by $|\kappa \pm \frac{1}{2} - \kappa^*|$ and $|\kappa - \kappa^* \mp \frac{1}{2}|$ respectively, if that of the original representation $\mathfrak{D}_{\kappa,\kappa^*}$ was $|\kappa - \kappa^*|$.

Let κ_0 and κ_0^* be any two complex numbers such that $2(\kappa_0 - \kappa_0^*)$ is an integer. Consider all numbers of the forms $\kappa = \kappa_0 + \frac{1}{2}n$, $\kappa^* = \kappa_0^* + \frac{1}{2}n^*$, where n, n^* are integers, and construct the matrices $\mathfrak{U}^\alpha, \mathfrak{V}^\alpha, \mathfrak{U}^{\dot\beta}, \mathfrak{V}^{\dot\beta}$ as follows

$$\left.\begin{aligned}
(\kappa, \kappa^* \,|\, \mathfrak{U}^\alpha \,|\, \kappa - \tfrac{1}{2}, \kappa^*) &= \mathfrak{U}^\alpha(\kappa, \kappa^*),\\
(\kappa - \tfrac{1}{2}, \kappa^* \,|\, \mathfrak{V}^\alpha \,|\, \kappa, \kappa^*) &= \mathfrak{V}^\alpha(\kappa, \kappa^*),\\
(\kappa, \kappa^* \,|\, \mathfrak{U}^{\dot\alpha} \,|\, \kappa, \kappa^* - \tfrac{1}{2}) &= \mathfrak{U}^{\dot\alpha}(\kappa, \kappa^*),\\
(\kappa, \kappa^* - \tfrac{1}{2} \,|\, \mathfrak{V}^{\dot\alpha} \,|\, \kappa, \kappa^*) &= \mathfrak{V}^{\dot\alpha}(\kappa, \kappa^*),
\end{aligned}\right\} \tag{64}$$

all the other matrix elements being zero. Then it is easy to verify directly from (44), (52a) and (56) that $\mathfrak{U}^\alpha, \mathfrak{V}^\alpha$ commute with $\mathfrak{U}^{\dot\beta}, \mathfrak{V}^{\dot\beta}$, so that, for example,

$$\mathfrak{U}^\alpha(\kappa, \kappa^*)\, \mathfrak{V}^{\dot\beta}(\kappa - \tfrac{1}{2}, \kappa^* + \tfrac{1}{2}) = \mathfrak{V}^{\dot\beta}(\kappa, \kappa^* + \tfrac{1}{2})\, \mathfrak{U}^\alpha(\kappa^* + \tfrac{1}{2}, \kappa).$$

Similarly define $J^{\alpha\beta}$ and $J^{\dot\alpha\dot\beta}$ by

$$J^{\alpha\beta} = \mathfrak{U}^\alpha \mathfrak{V}^\beta + \mathfrak{U}^\beta \mathfrak{V}^\alpha, \tag{65a}$$

$$J^{\dot\alpha\dot\beta} = \mathfrak{U}^{\dot\alpha} \mathfrak{V}^{\dot\beta} + \mathfrak{U}^{\dot\beta} \mathfrak{V}^{\dot\alpha}. \tag{65b}$$

$(\kappa, \kappa^* \mid J^{\alpha\beta} \mid \kappa, \kappa^*)$ and $(\kappa, \kappa^* \mid J^{\dot\alpha\dot\beta} \mid \kappa, \kappa^*)$ are equal to the $I^{\alpha\beta}$ and $I^{\dot\alpha\dot\beta}$ of the representation $\mathfrak{D}_{\kappa,\kappa^*}$. Now, in order that an operator C_α may transform as an undotted spinor under Lorentz transformation, it is necessary and sufficient that

$$[J_{\alpha\beta}, C_\gamma] = \tfrac{1}{2} \underset{(\alpha\beta)}{S}\, \epsilon_{\gamma\alpha} C_\beta, \tag{66a}$$

$$[J_{\dot\alpha\dot\beta}, C_\gamma] = 0. \tag{66b}$$

A similar condition holds with the interchange of dotted and undotted indices for a dotted spinor $C_{\dot\gamma}$. The commutation relations of $\mathfrak{U}^\alpha$, $\mathfrak{V}^\beta$ and $\mathfrak{U}^{\dot\alpha}$, $\mathfrak{V}^{\dot\beta}$ thus ensure that the former transform as undotted spinors and the latter as dotted spinors.

4. UNITARY REPRESENTATIONS

Since I is real for unitary representations it follows from (41) that either $\kappa^* + \kappa + 1$ is pure imaginary or else $\kappa^* - \kappa = \pm j = 0$. In the first case (20) is satisfied automatically and κ, κ^* are given by (1*a*). In the second case it follows from (20) that $\kappa = \kappa^*$ is real and

$$1 \geqslant (\kappa + \kappa^* + 1)^2 = (2\kappa + 1)^2,$$

so that $\frac{1}{2} \geqslant |\kappa + \frac{1}{2}|$, i.e. $0 \geqslant \kappa = \kappa^* \geqslant -1$ in accordance with (1*b*). In particular, if κ, κ^* are both of the form $\frac{1}{4}n\,(n \neq -2)$ then $j^+ = |\kappa + \frac{1}{2}| + |\kappa^* + \frac{1}{2}|$ and

$$\pm j^- = |\kappa + \tfrac{1}{2}| - |\kappa^* + \tfrac{1}{2}|.$$

Therefore $\mathfrak{D}^-_{\kappa,\kappa^*}$ is unitary only if $\frac{1}{2} \geqslant |\kappa + \frac{1}{2}| = |\kappa^* + \frac{1}{2}|$ while $\mathfrak{D}^+_{\kappa,\kappa^*}$ is unitary whenever $|\kappa + \frac{1}{2}| = |\kappa^* + \frac{1}{2}|$, since in the latter case

$$\pm iI = (|\kappa + \tfrac{1}{2}| - |\kappa^* + \tfrac{1}{2}|)(|\kappa + \tfrac{1}{2}| + |\kappa^* + \tfrac{1}{2}|) = 0$$

and $j = j^+ \neq 0$. By reflexion I goes over into $-I$ so that a representation is reflexion invariant only if I is zero, i.e. $\kappa + \frac{1}{2} = \pm(\kappa^* + \frac{1}{2})$. Hence the representations corresponding to (1*b*) are always reflexion invariant while those corresponding to (1*a*) are so only if $\nu = 0$.

5. EXPINORS

Consider two complex variables u_1 and u_2 which transform as components of a spinor under Lorentz transformations. Put

$$u_{\dot 1} = \overline{u_1}, \quad u_{\dot 2} = \overline{u_2}, \tag{67a}$$

$$\dot u_1 = -\overline{u_2}, \quad \dot u_2 = \overline{u_1}, \tag{67b}$$

where the bar denotes complex conjugate. Then $u_{\dot 1}, u_{\dot 2}$ transform as components of a dotted spinor and $\dot u_1, \dot u_2$ transform the same way as u_1, u_2 under spatial rotations, i.e. under unitary spinor transformations (see Van der Waerden 1932). Now consider terms of the type

$$\frac{u_1^p u_2^q u_{\dot 1}^r u_{\dot 2}^s}{\Delta^{n-\lambda}} = \frac{(-1)^s u_1^p u_2^q \dot u_1^s \dot u_2^r}{\Delta^{n-\lambda}}, \tag{68}$$

where $$\Delta = u_1 u_{\dot 1} + u_2 u_{\dot 2} = u_1 \dot u_2 - u_2 \dot u_1 = |u_1|^2 + |u_2|^2 \tag{69}$$

and p, q, r, s, n are integers $\geqslant 0$ while λ is a fixed complex number. Notice that on account of (69) the various terms for different values of p, q, r, s, n are in general *not* linearly independent. To remedy this defect one proceeds as follows. Put

$$U_M^{K,j} = \sum_{M=m+m'} \frac{2k!}{k-m!\,k+m!}\,\frac{2k'!}{k'-m'!\,k'+m'!} u_1^{k-m} u_2^{k+m} \dot{u}_1^{k'-m'} \dot{u}_2^{k'+m'}, \tag{70a}$$

$$K = k+k', \tag{70b}$$

$$j = k'-k, \tag{70c}$$

where $2k, 2k'$ are integers $\geqslant 0$. M can take the values $K, K-1, \ldots, -K$ and similarly m and m' the values $k, k-1, \ldots, -k$ and $k', k'-1, \ldots, -k'$ respectively. The summation in (70a) is over all values of m and m' such that $m+m' = M$. It is well known (see Van der Waerden 1932) that every polynomial $P^{k,k'}$ which is homogeneous in each of the two sets of variables u and $\dot{u}$ and has the degrees $2k$ and $2k'$ respectively can be expressed *uniquely* in terms of $U_M^{K,j}$ as follows:

$$P^{k,k'} = \sum_{0 \leqslant \nu \leqslant 2k,\, 2k'} \sum_{M=-K+\nu}^{K-\nu} A_{M,\nu} \Delta^{\nu} U_M^{K-\nu,j}. \tag{71}$$

ν runs through all integral values in the range $2k, 2k' \geqslant \nu \geqslant 0$ and $A_{M,\nu}$ are constants. Since every polynomial in u and $\dot{u}$ can be split up into a sum of homogeneous polynomials the expansion (71) is always possible and is *unique*. One can therefore express the numerator of (68) in the form (71) and arrange the result according to descending powers of Δ. Now consider a series of terms of the type (68) which is homogeneous and of degree $\lambda - j$ and $\lambda + j$ in u and $\dot{u}$ respectively, where

$$2j = r+s-p-q$$

is an integer. Then it follows from (71) that it can be put in the form

$$\sum_{K \geqslant |j|}^{\infty} \sum_{M=-K}^{K} a_M^{K,j} \frac{U_M^{K,j}}{\Delta^{K-\lambda}}, \tag{72}$$

where $a_M^{K,j}$ are constants. K runs through the values $|j|, |j|+1, \ldots, \infty$. Now apply a spinor transformation to u_1 and u_2

$$u_\mu = \alpha_\mu{}^\nu u'_\nu, \quad u_{\dot{\mu}} = \overline{\alpha_\mu{}^\nu} u'_{\dot{\nu}}, \tag{73}$$

and as before put

$$\dot{u}'_1 = -\overline{u'_2}, \quad \dot{u}'_2 = \overline{u'_1}. \tag{74}$$

Then $U_M^{K,j} \equiv U_M^{K,j}(u, \dot{u})$ goes over into a polynomial in u' and $\dot{u}'$ of degree $K-j$ and $K+j$ respectively, which from (71) can again be expressed linearly in terms of $\Delta'^{\nu} U'^{K-\nu,j}_M$, where $\Delta' = |u'_1|^2 + |u'_2|^2$ and $U'^{K,j}_M = U_M^{K,j}(u', \dot{u}')$. Further

$$\begin{aligned} \Delta &= |u_1|^2 + |u_2|^2 = |\alpha_1{}^1 u'_1 + \alpha_1{}^2 u'_2|^2 + |\alpha_2{}^1 u'_1 + \alpha_2{}^2 u'_1|^2 \\ &= \|\alpha\| (|u'_1|^2 + |u'_2|^2) - (|\alpha_1{}^1 \dot{u}'_1 + \alpha_1{}^2 \dot{u}'_1|^2 + |\alpha_2{}^1 \dot{u}'_1 + \alpha_2{}^2 \dot{u}'_2|^2), \end{aligned} \tag{75}$$

where $\|\alpha\| \equiv |\alpha_1{}^1|^2 + |\alpha_1{}^2|^2 + |\alpha_2{}^1|^2 + |\alpha_2{}^2|^2 > 0$. Thus

$$\Delta = \|\alpha\|\Delta'\left\{1 - \frac{D'}{\|\alpha\|\Delta'}\right\}, \tag{76}$$

where
$$D' = |\alpha_1{}^1\dot{u}'_1 + \alpha_1{}^2\dot{u}'_2|^2 + |\alpha_2{}^1\dot{u}'_1 + \alpha_2{}^2\dot{u}'_2|^2 \tag{77}$$

and
$$0 < \frac{D'}{\|\alpha\|\Delta'} < 1, \tag{78}$$

since $\Delta, \Delta' > 0$, if one excludes the case $u_1 = u_2 = 0$. Therefore one can expand $1/\Delta^{K-\lambda}$ in (72) by the binomial theorem in ascending powers of $D'/\|\alpha\|\Delta'$. Now D' is a polynomial in u' and $\dot{u}'$. Hence the product of its powers with $U'^{K,j}_M$ can again be expressed linearly in the form (71). Ignoring for the moment the question of convergence, which will be examined presently one may collect the coefficients of $U'^{K,j}_M/\Delta'^{K-\lambda}$ together and write

$$\sum_{K,M}^{\infty} a^{K,j}_M \frac{U^{K,j}_M}{\Delta^{K-\lambda}} = \sum_{K,M}^{\infty} a'^{K,j}_M \frac{U'^{K,j}_M}{\Delta'^{K-\lambda}}, \tag{79}$$

where $a'^{K,j}_M$ are the new coefficients. The numbers $\lambda - j$ and $\lambda + j$ being the degrees of the series in the variables u and $\dot{u}$ respectively, are clearly left unaltered by the homogeneous linear transformation (73). Hence only terms of the type $U'^{K,j}_M/\Delta'^{K-\lambda}$ can occur on the right-hand side of (79). The linear transformation $a^{K,j}_M \rightarrow a'^{K,j}_M$ of the coefficients thus constitutes a representation of the Lorentz transformation associated with (73).

To examine rigorously the validity of our considerations make the substitution

$$u_1 = \rho\cos\tfrac{1}{2}\theta\, e^{\frac{1}{2}i\chi} e^{\frac{1}{2}i\phi}, \quad u_{\dot{1}} = \rho\cos\tfrac{1}{2}\theta\, e^{-\frac{1}{2}i\chi} e^{-\frac{1}{2}i\phi}, \tag{80a}$$

$$u_2 = \rho\sin\tfrac{1}{2}\theta\, e^{\frac{1}{2}i\chi} e^{-\frac{1}{2}i\phi}, \quad u_{\dot{2}} = \rho\sin\tfrac{1}{2}\theta\, e^{-\frac{1}{2}i\chi} e^{+\frac{1}{2}i\phi}, \tag{80b}$$

where ρ, θ, χ, ϕ are real and $\rho > 0, \pi \geqslant \theta \geqslant 0,\ 2\pi > \phi \geqslant 0,\ 4\pi > \chi \geqslant 0$. Put $x = \cos\theta$ and

$$\frac{1}{\sqrt{(2\pi)}} A^{K,j}_M = \sqrt{\left(\frac{K-j!\,K+j!}{K-M!\,K+M!}\frac{2}{2K+1}\right)} a^{K,j}_M. \tag{81a}$$

Then (72) becomes

$$\rho^{2\lambda} e^{-ij\chi} \sum_{K\geqslant|j|}^{\infty} \sum_{M=-K}^{K} A^{K,j}_M \frac{1}{\sqrt{(2\pi)}} P^{K,j}_M(x)\, e^{-iM\phi}, \tag{81b}$$

where
$$P^{K,j}_M(x) = \frac{(-1)^{K+j}}{2^K} \sqrt{\left(\frac{K-M!}{K+M!}\frac{(K+\frac{1}{2})}{K+j!\,K-j!}\right)}$$
$$\times (1-x)^{\frac{1}{2}(M-j)}(1+x)^{\frac{1}{2}(M+j)} \frac{d^{K+M}}{dx^{K+M}}\{(1-x)^{K+j}(1+x)^{K-j}\}. \tag{81c}$$

The functions $Q^{K,j}_M \equiv \dfrac{1}{\sqrt{(2\pi)}} P^{K,j}_M(x)\, e^{-iM\phi}$ are orthogonal and normalized, i.e.

$$\int_{\phi=0}^{2\pi}\int_{x=-1}^{1} \overline{Q^{K',j}_{M'}}\, Q^{K,j}_M\, dx\, d\phi = \begin{cases} 1 & \text{if} \quad K = K', M = M', \\ 0 & \text{otherwise.} \end{cases} \tag{82}$$

For convenience arrange these orthogonal functions $Q_M^{K,j}$ in a sequence

$$Q_\nu \, (\nu = 1, 2, \dots, \infty)$$

and denote the interval $-1 \leqslant x \leqslant 1, 0 \leqslant \phi \leqslant 2\pi$ by G. Then (82) can be written as

$$\int_G \overline{Q_\mu} Q_\nu \, dx \, d\phi = \delta_{\mu\nu}. \tag{83}$$

Denote a series $\sum_1^\infty A_\nu Q_\nu$ by A. A_ν will be called the coefficients of A. It is important to note that A is not the sum of the series since this sum may not exist in general. A is merely an abbreviation for the infinite sequence of complex numbers A_ν $(\nu = 1, 2, \dots, \infty)$ which completely characterizes the series. Two series A and B are to be regarded as distinct unless $A_\nu = B_\nu$ for all ν. If c is a constant cA denotes a series with the coefficients cA_ν. Similarly $A + B$ stands for the series with the coefficients $A_\nu + B_\nu$. $A = 0$ is the series such that $A_\nu = 0$ for all ν. The norm $\| A \|$ of a series A is a non-negative number defined by

$$\| A \|^2 = \sum_1^\infty | A |^2. \tag{84}$$

Clearly $\| A \| > 0$ unless $A = 0$ when $\| A \| = 0$. If the series on the right-hand side of (84) is divergent we say the norm is infinite.

Let $f(x, \phi)$ be a continuous function of x and ϕ defined in G. Then by the norm of f is meant the number $\| f \| \geqslant 0$ given by

$$\| f \|^2 = \int_G | f |^2 \, dx \, d\phi. \tag{85}$$

Now denote by A^N the partial sum $\sum_1^N A_\nu Q_\nu$ of the series A. A^N is a continuous function of x and ϕ. Hence from (85) and (83)

$$\| A^N \|^2 = \sum_1^N | A_\nu |^2, \tag{86a}$$

so that

$$\| A \|^2 = \lim_{N \to \infty} \| A^N \|^2. \tag{86b}$$

This rather natural connexion justifies the use of the same term norm for both A and f. Now it is well known (see Courant & Hilbert 1931) that the functions Q_ν form a complete set in the sense that for every continuous function defined in G there exists a series A such that

$$\lim_{N \to \infty} \| f - A^N \| = 0. \tag{87}$$

Clearly this series is unique, since if A and B be two such series then by Schwarz's inequality which clearly holds for norms

$$\| A - B \| = \lim_{N \to \infty} \| A^N - B^N \| \leqslant \lim_{N \to \infty} \{ \| f - A^N \| + \| f - B^N \| \} = 0.$$

Hence $A = B$. It can be proved from (87) and (83) (see Courant & Hilbert 1931) that

$$A_\nu = \int_G \overline{Q_\nu} f \, dx \, d\phi = (Q_\nu, f) \tag{88a}$$

and

$$\|f\| = \|A\|, \tag{88b}$$

so that A has a finite norm. Now consider (81*b*). Write it as

$$\rho^{2\lambda} e^{-ij\chi} A. \tag{89}$$

Apply the spinor transformation to the u's and introduce the four real variables $\rho', \chi', \theta', \phi'$ corresponding to u_1', u_2' by equations similar to (80). Put $x' = \cos\theta'$. Then ρ, χ, x, ϕ are functions of ρ', χ', x', ϕ'. Notice that

$$\frac{\rho^2}{\rho'^2} = \frac{|u_1|^2 + |u_2|^2}{|u_1'|^2 + |u_2'|^2}, \quad e^{i(\chi-\chi')} = \frac{u_1 u_2}{u_1' u_2'} \left| \frac{u_1' u_2'}{u_1 u_2} \right|, \tag{90}$$

and so ρ^2/ρ'^2 and $e^{i(\chi-\chi')}$ are functions of x' and ϕ' alone. Obviously the same is true of x and ϕ. Therefore

$$\begin{aligned} \rho^{2\lambda} e^{-ij\chi} A^N &= \rho'^{2\lambda} e^{-ij\chi'} \left\{ \left(\frac{\rho^2}{\rho'^2} \right)^\lambda e^{-ij(\chi-\chi')} \sum_1^N A_\nu Q_\nu(x, \phi) \right\} \\ &\equiv \rho'^{2\lambda} e^{-ij\chi'} f_N(x', \phi'). \end{aligned} \tag{91}$$

Clearly $f_N(x', \phi')$ is a continuous function of x', ϕ'. Now put

$$A'_{\nu,N} = \int_G \overline{Q_\nu(x', \phi')} f_N(x', \phi') \, dx' \, d\phi', \tag{92}$$

and denote by A'_N the series whose coefficients are $A'_{\nu,N}$. From the completeness of the set Q_ν it follows that

$$\lim_{n\to\infty} \|f_N - A'^n_N\|^2 = \lim_{n\to\infty} \int_G |f_N(x', \phi') - A'^n_N(x', \phi')|^2 \, dx' \, d\phi' = 0, \tag{93}$$

where $A'^n_N(x', \phi') = \sum_{\nu=1}^n A'_{\nu,N} Q_\nu(x', \phi')$. Therefore

$$\begin{aligned} \|A'_{N+p} - A'_N\| &= \lim_{n\to\infty} \|A'^n_{N+p} - A'^n_N\| \\ &\leqslant \lim_{n\to\infty} \{\|f_{N+p} - A'^n_{N+p}\| + \|f_N - A'^n_N\| + \|f_{N+p} - f_N\|\} \\ &= \|f_{N+p} - f_N\| \end{aligned} \tag{94}$$

from (93). But from (91)

$$\|f_{N+p} - f_N\|^2 = \int_G \left| \left(\frac{\rho^2}{\rho'^2} \right)^\lambda \right|^2 \left| \sum_{N+1}^{N+p} A_\nu Q_\nu(x, \phi) \right|^2 dx' \, d\phi'. \tag{95}$$

Now since the determinant of any spinor transformation is unity it follows that

$$\frac{\partial(u_1, u_2, u_{\dot 1}, u_{\dot 2})}{\partial(\rho', \chi', \phi', \theta')} = \frac{\partial(u_1', u_2', u_{\dot 1}', u_{\dot 2}')}{\partial(\rho', \chi', \phi', \theta')} = \tfrac{1}{2} \rho'^3 \sin\theta'. \tag{96}$$

(Here the usual notation for the Jacobian has been used.) Therefore

$$\rho^3 \sin\theta \frac{\partial(\rho,\chi,\phi,\theta)}{\partial(\rho',\chi',\phi',\theta')} = \rho'^3 \sin\theta',$$

or

$$\frac{\partial(\rho,\chi,x,\phi)}{\partial(\rho',\chi',x',\phi')} = \frac{\rho'^3}{\rho^3}. \tag{97}$$

But since ρ/ρ', $e^{i(\chi-\chi')}$, x and ϕ are functions of x' and ϕ' alone

$$\frac{\partial(\rho,\chi,x,\phi)}{\partial(\rho',\chi',x',\phi')} = \frac{\rho}{\rho'}\frac{\partial(x,\phi)}{\partial(x',\phi')},$$

so that from (97)

$$\frac{\partial(x',\phi')}{\partial(x,\phi)} = \left(\frac{\rho}{\rho'}\right)^4 = \left(\frac{\Delta}{\Delta'}\right)^2. \tag{98}$$

Therefore, taking x, ϕ as the independent variables in (95), one gets

$$\|f_{N+p} - f_N\|^2 = \int_G \left|\left(\frac{\Delta}{\Delta'}\right)^{\lambda+1}\right|^2 \left|\sum_{N+1}^{N+p} A_\nu Q_\nu(x,\phi)\right|^2 dx\,d\phi.$$

But from (76)

$$\frac{\Delta}{\Delta'} < \|\alpha\|.$$

Similarly by considering the inverse transformation α^{-1} one finds that $\Delta'/\Delta < \|\alpha^{-1}\|$. Hence

$$\frac{1}{\|\alpha^{-1}\|} < \frac{\Delta}{\Delta'} < \|\alpha\|.$$

Therefore $|(\Delta/\Delta')^{\lambda+1}|^2$ is bounded for every fixed value of λ and α, and so

$$\|f_{N+p} - f_N\|^2 \leqslant C(\alpha,\lambda) \int_G \left|\sum_{N+1}^{N+p} A_\nu Q_\nu(x,\phi)\right|^2 dx\,d\phi, \tag{99}$$

where $C(\alpha,\lambda)$ is a real positive constant depending on α and λ. Hence from (94) and (99)

$$\|A'_{N'} - A'_N\| \leqslant C(\alpha,\lambda) \sum_{N_0}^{\infty} |A_\nu|^2 \quad (N, N' \geqslant N_0).$$

So that from the convergence of $\sum_1^\infty |A_\nu|^2$ it follows that

$$\lim_{\substack{N'\to\infty\\ N\to\infty}} \|A'_{N'} - A'_N\| = 0. \tag{100}$$

Therefore $A'_{\nu,N}$ tends to a limit A'_ν as $N \to \infty$ in such a way that if A' be the corresponding series then

$$\lim_{N\to\infty} \|A' - A'_N\| = 0. \tag{101}$$

Also clearly

$$\sum_{\nu=1}^{n} |A'_{N,\nu}|^2 \leqslant \int_G |f_N(x',\phi')|^2 dx'\,d\phi' = \int_G \left|\left(\frac{\Delta}{\Delta'}\right)^{\lambda+1}\right|^2 |A^N(x,\phi)|^2 dx\,d\phi$$
$$\leqslant C(\alpha,\lambda) \sum_{\nu=1}^{N} |A_\nu|^2 \leqslant C(\alpha,\lambda)\, \|A\|^2.$$

Therefore $\|A'\|$ is finite. A' has the property that

$$\lim_{\substack{N\to\infty\\N'\to\infty}} \int_G |\rho^{2\lambda}e^{-ij\chi}A^N(x,\phi)-\rho'^{2\lambda}e^{-ij\chi'}A'^{N'}(x',\phi')|^2\,dx'd\phi' = 0. \tag{102}$$

This follows from the inequality

$$\begin{aligned}\|f_N-A'^n\| &\leqslant \|f_{N_0}-A'^n_{N_0}\|+\|A'^n-A'^n_{N_0}\|+\|f_N-f_{N_0}\|\\ &\leqslant \|f_{N_0}-A'^n_{N_0}\|+\|A'-A'_{N_0}\|+\|f_N-f_{N_0}\|.\end{aligned} \tag{103}$$

By choosing n_0 and N_0 sufficiently large all the three terms on the right can be made arbitrarily small for $n \geqslant n_0, N \geqslant N_0$. On going over to the variables x, ϕ in (102) and using (98) one finds that

$$\lim_{\substack{N\to\infty\\N'\to\infty}} \int_G |\rho^{2\lambda}e^{-ij\chi}A^N(x,\phi)-\rho'^{2\lambda}e^{-ij\chi'}A'^{N'}(x',\phi')|^2\,dx\,d\phi = 0. \tag{104}$$

Thus (102) implies (104) and vice versa. Also it is clear that A' is the only series satisfying (102), since if B' were another such series then

$$\|A'^n-B'^n\| \leqslant \|f_N-A'^n\|+\|f_N-B'^n\|,$$

so that on making N and n tend to infinity

$$\|A'-B'\| = \lim_{n\leftarrow\infty}\|A'^n-B'^n\| = 0.$$

Hence $A' = B'$. The transformation $A \to A'$ is thus uniquely determined by the spinor transformation α. Clearly it is a linear transformation in the sense that if $A \to A'$ and $B \to B'$ then $A+B \to A'+B'$ and $cA \to cA'$, where c is a complex number. In fact it is a linear transformation of the Hilbert space consisting of all sequences for which $\sum_1^\infty |A_\nu|^2$ is convergent. These transformations thus form a representation of the Lorentz group. Also notice that

$$\|A'_N\|^2 = \int_G |f_N(x',\phi')|^2\,dx'd\phi' = \int_G \left|\left(\frac{\Delta}{\Delta'}\right)^{\lambda+1}\right|^2 |A^N(x,\phi)|^2\,dx\,d\phi,$$

so that if $\lambda+1$ is pure imaginary or zero then

$$\|A'_N\|^2 = \|A^N\|^2.$$

Therefore $$\|A'\| = \|A\|,$$

since $$\lim_{N\to\infty} |\,\|A'\|-\|A'_N\|\,| \leqslant \lim_{N\to\infty}\|A'-A'_N\| = 0$$

from (101). Thus the norm of the 'vector' A of the Hilbert space is unchanged by the transformation. The representation is therefore unitary. Besides these unitary representations there are others corresponding to the case $j = 0$, λ real and $|\lambda+1| \leqslant 1$ (cf. (1b)), but they do not leave the norm $\|A\|$ invariant.

To obtain the same matrix representation as that given in § 2 one has to choose

$$i^K\sqrt{\left(\frac{K-\lambda-1!}{K+\lambda+1!}\right)}\,Q_M^{K,j}\,e^{-ij\chi}\rho^{2\lambda} \tag{105a}$$

as the basic vectors. Here $z! = \Gamma(z+1)$ for any complex number z. In the exceptional case where λ is an integer $K-\lambda-1!$ or $K+\lambda+1!$ may become infinite. Then one can make use of the relation

$$\Gamma(z)\,\Gamma(1-z) = \frac{\pi}{\sin z\pi},$$

and ignore constant factors (including $1/\sin\pi = \infty$). Thus one gets the alternative forms

$$\frac{Q_M^{K,j}e^{-ij\chi}\rho^{2\lambda}}{\sqrt{(\lambda+1+K!\,\lambda-K!)}}, \tag{105b}$$

and

$$\sqrt{(K-\lambda-1!\,-\lambda-2-K!)}\,Q_M^{K,j}\,e^{-ij\chi}\rho^{2\lambda}. \tag{105c}$$

($105a, b, c$) respectively are applicable to the three cases $|j| \geqslant |\lambda+1|$, $\lambda \geqslant |j|$ and $-\lambda \geqslant |j|+2$, λ being an integer. Then ($105b$) and ($105c$) are finite representations. ($105c$) corresponds to an expansion in ascending powers of $1/\Delta$ and ignoring all powers higher than $1/\Delta^{-(2\lambda+2)}$, i.e. calculating the expansion mod $1/\Delta^{-(2\lambda+2)}$.

The theory developed so far for functions of the two variables x, ϕ can easily be extended to those of the three variables x, ϕ, χ defined in the interval H given by $-1 \leqslant x \leqslant 1, 0 \leqslant \phi \leqslant 2\pi, 0 \leqslant \chi \leqslant 4\pi$. Put

$$S_M^{K,j} = \frac{1}{\sqrt{(4\pi)}}\,Q_M^{K,j}e^{-ij\chi}, \tag{106}$$

and denote a series $\sum\limits_{\nu=1}^{\infty} \mathfrak{A}_\nu S_\nu \equiv \sum\limits_{j=-\infty}^{\infty}\ \sum\limits_{K\geqslant|j|} A_M^{K,j}S_M^{K,j}$

by $\mathfrak{A}$. The summation for j is over all integral and half-integral values. Put

$$\|\mathfrak{A}\|^2 = \sum_{\nu=1}^{\infty}|\mathfrak{A}_\nu|^2.$$

Let $f(x, \phi, \chi)$ be a continuous function of x, ϕ, χ defined in the interval H. Put

$$\|f\|^2 = \int_H |f|^2\,dx\,d\phi\,d\chi.$$

Now it is well known that the set S_ν is complete and therefore for every continuous f there exists a series $\mathfrak{A}$ such that

$$\lim_{N\to\infty}\|f-\mathfrak{A}^N\| = 0, \tag{107a}$$

where $\mathfrak{A}^N = \sum\limits_{\nu=1}^{N}\mathfrak{A}_\nu S_\nu$. Also

$$\|f\|^2 = \|\mathfrak{A}\|^2 = \lim_{N\to\infty}\|\mathfrak{A}^N\|^2, \tag{107b}$$

$$\mathfrak{A}_\nu = (S_\nu, f), \tag{107c}$$

where the symbol (g,f) is defined for any two continuous functions g and f by the equation

$$(g,f) = \overline{(f,g)} = \int_H \bar{g} f \, dx \, d\phi \, d\chi. \tag{108}$$

Let $\mathfrak{A}$ be a series with finite norm and let f be a continuous function defined on H. Then f is bounded, i.e. there exists a constant C such that $|f| \leqslant C$. Clearly $f\mathfrak{A}^N$ is a continuous function. Hence there exists a series $\mathfrak{B}_N$ satisfying the equation

$$\lim_{n\to\infty} \| f\mathfrak{A}^N - \mathfrak{B}_N^n \| = 0.$$

Now $$\| \mathfrak{B}_N^n - \mathfrak{B}_{N_0}^n \| \leqslant \| f\mathfrak{A}^{N_0} - \mathfrak{B}_{N_0}^n \| + \| f\mathfrak{A}^N - \mathfrak{B}_N^n \| + \| f\mathfrak{A}^N - f\mathfrak{A}^{N_0} \|,$$

and so $$\| \mathfrak{B}_N - \mathfrak{B}_{N_0} \| \leqslant \| f\mathfrak{A}^N - f\mathfrak{A}^{N_0} \|.$$

i.e. $$\| \mathfrak{B}_N - \mathfrak{B}_{N_0} \| \leqslant C^2 \sum_{N_0}^{\infty} | \mathfrak{A}_\nu |^2 \quad (N \geqslant N_0).$$

Therefore $\mathfrak{B}_N$ tends to a limit $\mathfrak{B}$ as $N\to\infty$ and $\|\mathfrak{B}\|$ is finite, since

$$\| \mathfrak{B} \|^2 = \lim_{N\to\infty} \int_H | f\mathfrak{A}^N |^2 \, dx \, d\phi \, d\chi \leqslant C^2 \, \| \mathfrak{A} \|^2.$$

Hence we conclude that $$\mathfrak{B}_\nu = \lim_{N\to\infty} (S_\nu, f\mathfrak{A}^N). \tag{109}$$

Now let D be any one of the three operators $\partial/\partial x$, $\partial/\partial\phi$ or $\partial/\partial\chi$ and let $f(x,\phi,\chi)$ be a continuous function whose three partial differential coefficients exist and are continuous everywhere on H. Also assume that

$$f(x,\phi,0) = f(x,\phi,4\pi) \quad \text{and} \quad f(x,0,\chi) = f(x,2\pi,\chi \pm 2\pi),$$

where the sign in $\chi \pm 2\pi$ is determined from the fact that the resulting value must lie between 0 and 2π. Then it is clear by partial integration that

$$(S_\nu, Df) = -(D^* S_\nu, f), \tag{110a}$$

where D^*S_ν is defined by $$D^* S_\nu = \overline{D\bar{S}_\nu}. \tag{110b}$$

But it is well known that D^*S_ν can be expressed as a finite linear combination of the S_σ's, i.e.

$$D^* S_\nu = \sum_\sigma S_\sigma (S_\sigma, D^* S_\nu), \tag{111}$$

where (S_σ, D^*S_ν) is zero for all except a finite number of values of σ. Hence from (110a) and (111)

$$\begin{aligned}(S_\nu, Df) &= -\sum_\sigma \overline{(S_\sigma, D^* S_\nu)}\,(S_\sigma, f) \\ &= \sum_\sigma \overline{(D S_\sigma, S_\nu)}\,(S_\sigma, f) \\ &= \sum_\sigma (S_\nu, D S_\sigma)\,(S_\sigma, f). \end{aligned} \tag{112}$$

Since Df is a continuous function the series $\mathfrak{B}$ with $\mathfrak{B}_\nu = (S_\nu, Df)$ 'represents' Df, i.e.

$$\lim_{n\to\infty} \| Df - \mathfrak{B}^n \| = 0,$$

and therefore

$$\| \mathfrak{B} \|^2 = \int_H | Df |^2 \, dx\, d\phi\, d\chi$$

is finite. Now (112) shows that the coefficients $\mathfrak{B}_\nu$ can be obtained by differentiating the series $\mathfrak{A}$ representing f ($\mathfrak{A}_\sigma = (S_\sigma, f)$) term by term, expressing DS_σ as a linear combination of S_ν and collecting coefficients of S_ν. Thus if $\mathfrak{A}$ 'represents' a continuous function with continuous partial derivatives then the series $D\mathfrak{A}$ obtained by term-by-term differentiation represents Df and therefore $\| D\mathfrak{A} \|$ is finite.

Now from a series $\mathfrak{A}$ one can always go back to the series of the type (72) by making use of the relation

$$S_M^{K,j} = \frac{(-1)^{K+j}}{\sqrt{8\pi}} \sqrt{\left(\frac{K-M!\,K+M!}{K-j!\,K+j!}\,(K+\tfrac{1}{2})\right)}\,\frac{U_M^{K,j}}{\Delta^K}.$$

Formally put

$$\psi(u, \dot{u}) = \sum_{j=-\infty}^{\infty} \sum_{K \geqslant |j|}^{\infty} \sum_M a_M^{K,j} \frac{U_M^{\dot{K},j}}{\Delta^{K-\lambda}}, \tag{113a}$$

and write

$$\begin{aligned} \| \psi \|^2 &= 8\pi^2 \sum_{j=-\infty}^{\infty} \sum_{K \geqslant |j|}^{\infty} \sum_M \frac{K-j!\,K+j!}{K-M!\,K+M!} \frac{2}{2K+1} | a_M^{K,j} |^2 \\ &= \| \mathfrak{A} \|^2. \end{aligned} \tag{113b}$$

For brevity we say that ψ is a continuous function of u and $\dot{u}$ if there exists a continuous function $f(x, \phi, \chi)$ defined on H with

$$f(x, \phi, 0) = f(x, \phi, 4\pi) \quad \text{and} \quad f(x, 0, \chi) = f(x, 2\pi, \chi \pm 2\pi)$$

such that

$$\mathfrak{A}_\nu = (S_\nu, f).$$

Such a function, if it exists, is uniquely determined by $\mathfrak{A}$, since if f and f' were two such functions then

$$\| f - f' \| \leqslant \lim_{N\to\infty} \| f - \mathfrak{A}^N \| + \lim_{N\to\infty} \| f' - \mathfrak{A}^N \| = 0,$$

so that

$$\| f - f' \|^2 = \int_H | f - f' |^2 \, dx\, d\phi\, d\chi = 0.$$

But as f and f' are continuous this implies that $f = f'$ everywhere on H. Similarly, if f has got continuous partial derivatives everywhere on H we say ψ is continuously differentiable. Hence the above result can be expressed by saying that *if ψ is continuously differentiable it can be differentiated term by term.*

Notice that if ψ is continuous in one Lorentz frame, then it is continuous in every other. Since if $\mathfrak{A}$ is the corresponding series and $\mathfrak{A}'$ its transform in any other Lorentz frame then

$$\lim_{\substack{N\to\infty \\ N'\to\infty}} \int_H | \rho^{2\lambda} \mathfrak{A}^N(x, \phi, \chi) - \rho'^{2\lambda} \mathfrak{A}'^{N'}(x', \phi', \chi') |^2 \, dx'\, d\phi'\, d\chi' \tag{114a}$$

and

$$\lim_{N\to\infty} \int_H | \rho^{2\lambda} \mathfrak{A}^N - \rho^{2\lambda} f |^2 \, dx\, d\phi\, d\chi = 0. \tag{114b}$$

where f is continuous. Define

$$f'(x',\phi',\chi') = \left(\frac{\rho}{\rho'}\right)^{2\lambda} f(x,\phi,\chi), \tag{114c}$$

then clearly $f'(x,\phi,\chi)$ is continuous, and as is easily proved

$$\lim_{N\to\infty} \|f' - \mathfrak{A}'^N\| = 0, \tag{114d}$$

Therefore the transform $\psi'(u,\dot u)$ is continuous. The same is true for the property of continuous differentiability. In particular, if $\psi(u,\dot u)$ is such that in some particular Lorentz frame it has only a finite number of terms in the series, i.e. $\mathfrak{A}_\nu = 0$ for all $\nu > N$ then clearly $\psi(u,\dot u)$ is continuously differentiable an arbitrary number of times.

One can now introduce the operators u_μ, $\partial/\partial u_\mu$, $u_{\dot\mu}$ and $\partial/\partial u_{\dot\mu}$. $\partial/\partial u_\mu$ is defined by

$$\frac{\partial}{\partial u_\mu} = \frac{\partial\rho}{\partial u_\mu}\frac{\partial}{\partial\rho} + \frac{\partial\chi}{\partial u_\mu}\frac{\partial}{\partial\chi} + \frac{\partial x}{\partial u_\mu}\frac{\partial}{\partial x} + \frac{\partial\phi}{\partial u_\mu}\frac{\partial}{\partial\phi},$$

where $\partial\xi/\partial u_\mu$ is $\left\{\dfrac{\partial(u_1,u_2,u_{\dot1},u_{\dot2})}{\partial(\rho,\chi,x,\phi)}\right\}^{-1}$ times the minor of $\partial u_\mu/\partial\xi$ in the Jacobian

$$\frac{\partial(u_1,u_2,u_{\dot1},u_{\dot2})}{\partial(\rho,\chi,x,\phi)} \quad (\xi = \rho,\chi,x \text{ or } \phi).$$

$\partial/\partial u_{\dot\mu}$ is defined similarly. u_μ, $u_{\dot\mu}$, $\partial\xi/\partial u_\mu$ and $\partial\xi/\partial u_{\dot\mu}$ are clearly continuous functions of x,ϕ,χ (apart from a possible factor ρ or $1/\rho$) and therefore from (109) and (112) the operators u_μ, $u_{\dot\mu}$, $\partial/\partial u_\mu$ and $\partial/\partial u_{\dot\mu}$ can be applied to every continuously differentiable series ψ. In this term by term application $\partial/\partial u_1$ corresponds formally to partial differentiation with respect to u_1 keeping u_2, $u_{\dot1}$, $u_{\dot2}$ constant. Similarly for $\partial/\partial u_2$, $\partial/\partial u_{\dot1}$ and $\partial/\partial u_{\dot2}$.

Consider an infinitesimal spinor transformation

$$u_\mu = u'_\mu + \eta_\mu{}^\nu u'_\nu, \tag{115a}$$

$$u_{\dot\mu} = u'_{\dot\mu} + \eta_{\dot\mu}{}^{\dot\nu} u'_{\dot\nu}. \tag{115b}$$

$(\eta_{\dot\mu}{}^{\dot\nu} = \overline{\eta_\mu{}^\nu})$ applied to the expinor space. From the condition that the determinant of the transformation should be 1 so that the fundamental spinor $\epsilon_{\mu\nu}$ remains invariant it follows that

$$\eta_{\mu\nu} = \eta_{\nu\mu}. \tag{115c}$$

Let $\psi(u,\dot u)$ be a series which is homogeneous and continuously differentiable so that $\psi = \rho^{2\lambda}\mathfrak{A}$. Under (115) ψ goes over to $\psi' = \rho^{2\lambda}\mathfrak{A}'$ where $\mathfrak{A}'$ is defined by (114a). Since ψ is continuously differentiable there exist continuously differentiable functions $f(x,\phi,\chi)$ and $f'(x',\phi',\chi')$ satisfying (114b,c,d). Write $\rho^{2\lambda}f(x,\phi,\chi) = \Psi(u,\dot u)$ and $\rho'^{2\lambda}f'(x',\phi',\chi') = \Psi'(u',\dot u')$. Then

$$\Psi'(u',\dot u') = \Psi(u,\dot u). \tag{116}$$

From (116), (115) and the definition of $\partial^{\mu} \equiv \partial/\partial u_{\mu}$ and $\partial^{\dot\mu} \equiv \partial/\partial u_{\dot\mu}$, it follows that

$$\begin{aligned}\Psi'(u,\dot u) &= \Psi'(u',\dot u') + \eta_{\mu}{}^{\nu} u'_{\nu} \frac{\partial}{\partial u'_{\mu}} \Psi' + \eta_{\dot\mu}{}^{\dot\nu} u'_{\dot\nu} \frac{\partial}{\partial u'_{\dot\mu}} \Psi' \\ &= \Psi(u,\dot u) + \eta_{\mu}{}^{\nu} u_{\nu} \partial^{\mu} \Psi + \eta_{\dot\mu}{}^{\dot\nu} u_{\dot\nu} \partial^{\dot\mu} \Psi \\ &= \{1 - \eta_{\mu\nu} I^{\mu\nu} - \eta_{\dot\mu\dot\nu} I^{\dot\mu\dot\nu}\} \Psi(u,\dot u) \end{aligned} \tag{117}$$

up to quantities of the first order. Here

$$I^{\mu\nu} = \tfrac{1}{2}(u^{\mu}\partial^{\nu} + u^{\nu}\partial^{\mu}), \tag{118a}$$

$$I^{\dot\mu\dot\nu} = \tfrac{1}{2}(u^{\dot\mu}\partial^{\dot\nu} + u^{\dot\nu}\partial^{\dot\mu}). \tag{118b}$$

Since ψ is continuously differentiable the functions $I^{\mu\nu}\Psi(u,\dot u)$ and $I^{\dot\mu\dot\nu}\Psi(u,\dot u)$ are represented by the series $I^{\mu\nu}\psi$ and $I^{\dot\mu\dot\nu}\psi$ respectively.

Now the Lorentz transformation

$$x_k = x'_k + \epsilon_k{}^l x'_l, \quad \epsilon_{kl} = -\epsilon_{lk} \tag{119a}$$

associated to (115) is given by

$$\epsilon^{kl} = -\tfrac{1}{2}\eta_{\alpha}{}^{\beta}\, \sigma^{k}_{\beta\dot\lambda}\sigma^{l,\alpha\dot\lambda} - \tfrac{1}{2}\eta_{\dot\lambda}{}^{\dot\mu}\sigma^{k}_{\alpha\dot\mu}\sigma^{l,\alpha\dot\lambda}, \tag{119b}$$

where the σ-symbols have their usual meaning (see Van der Waerden 1932). (119b) is obtained immediately from the condition that a spinor transformation applied together with the associated Lorentz transformation leaves $\sigma^{k}_{\alpha\dot\lambda}$ unaltered. From (3) the transformation operator corresponding to (119a) is $1 - \frac{1}{2}\epsilon_{kl}I^{kl}$. Hence the connexion between $I^{\mu\nu}$, $I^{\dot\mu\dot\nu}$ and I_{kl} is given by

$$I_{\alpha}{}^{\beta} = \tfrac{1}{4} I_{kl}\sigma^{k}_{\alpha\dot\lambda}\sigma^{l,\dot\lambda\beta}, \quad I_{\dot\lambda}{}^{\dot\mu} = \tfrac{1}{4} I_{kl}\sigma^{k}_{\dot\lambda\alpha}\sigma^{\lambda,\alpha\dot\mu}. \tag{120}$$

On using the usual representation for $\sigma^{k}_{\alpha\dot\lambda}$ it is easily proved that $I_{\alpha\beta}$ and $I_{\dot\lambda\dot\mu}$ so obtained are the same as those introduced in § 3.

Notice that

$$[\partial^{\mu}, u^{\nu}] = \epsilon^{\nu\mu} = [u^{\mu}, \partial^{\nu}], \tag{121a}$$

$$[\partial^{\dot\mu}, u^{\dot\nu}] = \epsilon^{\dot\nu\dot\mu} = [u^{\dot\mu}, \partial^{\dot\nu}], \tag{121b}$$

and the dotted operators commute with the undotted ones. Also if ψ is homogeneous in the dotted and undotted pair of variables separately and of degree $2\kappa^* = \lambda + j$ and $2\kappa = \lambda - j$ respectively, then it follows immediately on term-by-term differentiation that

$$u_{\mu}\partial^{\mu}\psi = 2\kappa\psi, \tag{122a}$$

$$u_{\dot\mu}\partial^{\dot\mu}\psi = 2\kappa^*\psi. \tag{122b}$$

From (118), (121) and (122) it is clear that the properties of u_{μ}, ∂_{μ} are the same as those of $\mathfrak{U}_{\mu}$, $\mathfrak{B}_{\mu}$ respectively as defined in (64). In fact the matrix representation of u_{μ}, ∂_{μ} is identical with that given for $\mathfrak{U}_{\mu}$, $\mathfrak{B}_{\mu}$ in § 3 if one chooses either of the three functions (105a), (105b) or (105c) for all values of λ and j as basic vectors. Similarly $u_{\dot\mu} = \mathfrak{U}_{\dot\mu}$ and $\partial_{\dot\mu} = \mathfrak{B}_{\dot\mu}$.

The meaning of (43) and (45) is now obvious in terms of u_α and ∂_α. Let ψ_α be a vector of the space $\mathfrak{R} \times \mathfrak{R}_{\frac{1}{2}}$ such that the corresponding pair of series $\psi_\alpha(u, \dot{u})$ are continuously differentiable and are of degree 2κ and $2\kappa^*$ in u and $\dot{u}$ respectively. Then clearly $u^\alpha \psi_\alpha(u, \dot{u})$ and $\partial^\alpha \psi_\alpha(u, \dot{u})$ are two series of degrees $(2\kappa+1, 2\kappa^*)$ and $(2\kappa-1, 2\kappa^*)$ respectively and therefore transform according to the representation $\mathfrak{D}_{\kappa+\frac{1}{2}}, \kappa^*$ and $\mathfrak{D}_{\kappa-\frac{1}{2}}, \kappa^*$.

Also notice that the commutation relation (121a) remains unchanged under the formal substitution

$$u_\alpha \rightarrow \partial_\alpha, \quad \partial_\alpha \rightarrow u_\alpha.$$

By the same substitution $2\kappa = u_\alpha \partial^\alpha$ goes over into $\partial_\alpha u^\alpha = -(2\kappa+2)$, i.e. κ goes over into $-\kappa-1$. This corresponds to (59).

6. Physical applications

The irreducible representations of the Lorentz group discussed above may be used to describe the transformation properties of the wave function of a particle. Let ψ_α and $\psi_{\dot{\lambda}}$ be two vectors of the product spaces $\mathfrak{R} \times \mathfrak{R}_{\frac{1}{2}}$ and $\mathfrak{R} \times \mathfrak{R}_{\dot{\frac{1}{2}}}$ respectively. The corresponding series will be denoted by $\psi_\alpha(u, \dot{u})$ and $\psi_{\dot{\lambda}}(u, \dot{u})$, their coefficients being functions of the space-time coordinates x_k $(k = 0, 1, 2, 3)$. Assume that the functions Ψ_α and Ψ_λ associated to these series (cf. § 5) are continuous and differentiable a sufficient number of times with respect to u_μ, $u_{\dot{\mu}}$ and x_k. Then in analogy with the case of finite representations (Fierz 1939) one can set up an equation of motion of the particle as follows:

$$i\partial^{\alpha\dot{\lambda}}\psi_\alpha = \chi\psi^{\dot{\lambda}}, \tag{123a}$$

$$i\partial_{\alpha\dot{\lambda}}\psi^{\dot{\lambda}} = \chi\psi_\alpha, \tag{123b}$$

where

$$\partial_{\alpha\dot{\lambda}} = \sigma^k_{\alpha\dot{\lambda}}\partial_k, \quad \partial_k \equiv \frac{\partial}{\partial x^k}$$

and χ is a constant $\neq 0$. The second-order equation

$$\{\partial_k \partial^k + \chi^2\} \psi_\alpha = 0 \tag{124}$$

follows immediately. (123) can be put in a more elegant form by making use of the formalism of a recent paper (Harish-Chandra 1947). Introduce the 4×1 matrices T_α and $T_{\dot{\lambda}}$ (see Harish-Chandra 1946) and put

$$\Psi = T_\alpha \psi^\alpha + iT_{\dot{\lambda}}\psi^{\dot{\lambda}}. \tag{125}$$

Then there exist 1×4 matrices Γ^*_β and $\Gamma^*_{\dot{\mu}}$ having the following properties:

$$\Gamma^*_{\dot{\alpha}} T_\beta = 0 = \Gamma^*_\alpha T_{\dot{\mu}}, \tag{126a}$$

$$\Gamma^*_\alpha T_\beta = i\epsilon_{\alpha\beta}, \quad \Gamma^*_{\dot{\lambda}} T_{\dot{\mu}} = i\epsilon_{\dot{\lambda}\dot{\mu}}, \tag{126b}$$

$$iT_\alpha \Gamma^{*\alpha} + iT_{\dot{\alpha}}\Gamma^{*\dot{\alpha}} = 1, \tag{126c}$$

$$A_{\lambda\dot{\mu}} = 2(T_\lambda \Gamma_{\dot{\mu}} + T_{\dot{\mu}}\Gamma^*_\lambda), \tag{126d}$$

where $A_{\lambda\dot{\mu}} = \alpha_k \sigma^k_{\lambda\dot{\mu}}$, α_k being the usual 4×4 Dirac matrices satisfying the commutation rules

$$\alpha_k \alpha_l + \alpha_l \alpha_k = 2g_{kl}. \tag{127}$$

Therefore

$$\psi_\alpha = i\Gamma^*_\alpha \Psi, \quad \psi_{\dot{\mu}} = \Gamma^*_{\dot{\mu}} \Psi, \tag{128a}$$

and (123) is equivalent to

$$i\alpha^k \partial_k \Psi + \chi \Psi = 0. \tag{128b}$$

Ψ is a 4×1 matrix, each element of which is a series in $u, \dot{u}$ of degree $(2\kappa, 2\kappa^*)$ whose coefficients are functions of x_k.

Now choose a representation with $\lambda + 1 = 0$. Then it is unitary. For two given series $\psi = \rho^{2\lambda} e^{-i\vartheta\chi} A$ and $\phi = \rho^{2\lambda} e^{-i\vartheta\chi} B$ define the scalar product (ψ, ϕ) by

$$(\psi, \phi) = \lim_{N \to \infty} \int \overline{A^N} B^N dx\, d\phi = \sum_{\nu=1}^{\infty} \overline{A_\nu} B_\nu. \tag{129}$$

From Schwarz's inequality

$$\left| \sum_{\nu=N_0}^{N} \overline{A_\nu} B_\nu \right|^2 \leqslant \sum_{\nu=N_0}^{N} |A_\nu|^2 \sum_{\nu=N_0}^{N} |B_\nu|^2 \leqslant \|A\|^2 \|B\|^2,$$

and therefore (ψ, ϕ) exists and is finite. Also it is easy to show from (98) that (ψ, ϕ) is invariant under Lorentz transformations. Notice that

$$(\psi, \phi) = \overline{(\phi, \psi)} \quad \text{and} \quad (\psi, \psi) = \|\psi\|^2.$$

The charge and current density corresponding to (128b) can be given by

$$s_k = e\Psi^* \alpha_k \Psi, \tag{130}$$

where Ψ^* is defined by

$$\Psi^* = \tilde{\Psi} \Lambda, \tag{131a}$$

Λ being a non-singular Hermitian matrix such that

$$\alpha_k^\dagger = \Lambda \alpha_k \Lambda^{-1}. \tag{131b}$$

($\dagger$ denotes Hermitian conjugate.) $\tilde{\Psi}$ is the transposed of Ψ and the right side of (130) stands for

$$e \sum_{\rho,\sigma=1}^{4} (\alpha_k)_{\rho\sigma} (\Psi^*_\rho, \Psi_\sigma).$$

The charge density is therefore positive definite as in Dirac's theory of the electron. In fact (128b) can be derived from the Lagrangian

$$L = \frac{1}{2i} (\partial_k \Psi^* \alpha^k \Psi - \Psi^* \alpha^k \partial_k \Psi) + \chi \Psi^* \Psi$$

in the usual way. The analogy with the case of the electron is complete. Neither the energy density nor the total energy is positive, and therefore the theory can be quantized only in accordance with the exclusion principle. Write $\rho^{2\lambda} e^{-i\vartheta\chi} Q_\nu = V_\nu$ and $\Psi_\rho(u, \dot{u}) = \sum_{V=1}^{\infty} \Psi_{\rho,\nu} V_\nu$ and use a representation for α_k for which $\Lambda = \alpha_0$. Then in the quantized theory the coefficients $\Psi_{\rho,\nu}$ appear as operators satisfying the commutation rules

$$\Psi^\dagger_{\sigma,\mu}(\mathbf{x}, t) \Psi_{\rho,\nu}(\mathbf{x}', t) + \Psi_{\rho,\nu}(\mathbf{x}', t) \Psi^\dagger_{\sigma,\mu}(\mathbf{x}, t) = \delta_{\sigma\rho} \delta_{\nu\mu} \delta(\mathbf{x} - \mathbf{x}'). \tag{132}$$

Here $\mathbf{x}$ and t are the space and time coordinates and $\Psi^{\dagger}_{\sigma,\mu}$ is the Hermitian conjugate of $\Psi_{\sigma,\mu}$.

Notice that no subsidiary conditions are necessary to make the charge-density positive definite (cf. Fierz 1939). Further, this scheme describes particles of integral or half-integral spin according as j is half-integral or integral. This follows immediately on reducing the total representation space with respect to the rotation group. Hence particles with integral spin must be quantized here according to Fermi statistics. Such a state of affairs, however, is not possible for finite representations (Pauli 1940).

Now instead of (128*b*) one can equally well take the equation of motion of the particle to be

$$i\beta_k\partial^k\Psi+\chi\Psi = 0, \tag{133a}$$

where β_k are the Duffin-Kemmer matrices satisfying the commutation rules

$$\beta_k\beta_l\beta_m+\beta_m\beta_l\beta_k = g_{kl}\beta_m+g_{ml}\beta_k. \tag{133b}$$

In this case the energy density as calculated from the energy momentum tensor

$$\left.\begin{aligned} T_{kl} &= \Psi^*(g_{kl}-\beta_k\beta_l-\beta_l\beta_k)\,\Psi,\\ \Psi^* &= \tilde{\Psi}\Lambda,\\ \beta_k^{\dagger} &= \Lambda\beta_k\Lambda^{-1}, \quad \Lambda^{\dagger} = \Lambda \end{aligned}\right\} \tag{134}$$

is positive definite. The equation describes particles of integral or half-integral spin according as j is integral or half-integral. As usual it can be quantized according to Bose statistics through the commutation rules

$$[(\Psi^*\beta_0^2)_{\sigma,\mu}, (\beta_0\Psi')_{\rho,\nu}] = -(\beta_0^2)_{\rho\sigma}\delta_{\nu\mu}\delta(\mathbf{x}-\mathbf{x}'), \tag{135}$$

where $$(\Psi^*\beta_0^2)_{\sigma,\mu} = \sum_{\rho}\Psi^*_{\rho,\mu}(\beta_0^2)_{\rho,\sigma} \quad\text{and}\quad (\beta_0\Psi')_{\rho,\nu} = \sum_{\sigma}(\beta_0)_{\rho\sigma}\Psi'_{\sigma,\nu}.$$

Ψ' stands for $\Psi(\mathbf{x}',t)$. Electromagnetic interaction can be introduced in (128) and (133) in the usual way. Also it is clear that (128) and (133) are both reflexion invariant since the unitary representations with $\lambda+1 = 0$ are reflexion invariant.

So far no subsidiary conditions have been imposed on Ψ. As a result the particle thus described does not possess a definite spin in the sense that in the rest system the non-vanishing components of the wave function do not all necessarily transform according to one irreducible representation $\mathfrak{D}_s$ of the rotation group. The subsidiary conditions imposed by Dirac (1936) and Fierz (1939) on ψ_α and $\psi_{\dot\lambda}$ are as follows:

$$\partial^\alpha\psi_\alpha = \partial^{\dot\lambda}\psi_{\dot\lambda} = 0.$$

From (126) and (128*b*) they may be written as

$$\partial^\mu\Gamma^*_\mu\Psi = \partial^{\dot\mu}\Gamma^*_{\dot\mu}\Psi = 0, \tag{136a}$$

or otherwise as $$\partial^\lambda\partial^{\dot\mu}A_{\lambda\dot\mu}\Psi = 0. \tag{136b}$$

Then as shown by Fierz (1939) all non-vanishing components of Ψ in the rest system transform according to $\mathfrak{D}_s$ with $s = \kappa + \kappa^*$ (where κ, κ^* are both real, $\geqslant 0$ and of the form $\frac{1}{2}n$). One can, however, consider this question in a more general way at least in the unquantized force-free case. Consider the quantity $\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} I_{\alpha\beta} I_{\dot\lambda\dot\mu}\Psi$, where $I_{\alpha\beta}, I_{\dot\lambda\dot\mu}$ refer to the total representation space so that, for example, in the particular case of (128)

$$I_{\alpha\beta} = \tfrac{1}{2}i(T_\alpha \Gamma^*_\beta + T_\beta \Gamma^*_\alpha) + \tfrac{1}{2}(u_\alpha \partial_\beta + u_\beta \partial_\alpha).$$

In the rest system of the particle $\partial_k = 0$ $(k \neq 0)$ and $i\partial_0 = \pm\chi$ so that from (53 *a*), (52), (31 *a*) and (29)

$$\begin{aligned}
\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} I_{\alpha\beta} I_{\dot\lambda\dot\mu}\Psi &= -\chi^2 \sigma_0{}^{\alpha\dot\lambda}\sigma_0{}^{\beta\dot\mu} I_{\alpha\beta} I_{\dot\lambda\dot\mu}\Psi \\
&= -\chi^2\{I_{11}I_{\dot1\dot1} + 2I_{12}I_{\dot1\dot2} + I_{22}I_{\dot2\dot2}\}\Psi \\
&= -\tfrac{1}{4}\chi^2 (K_{\alpha\beta} + iL_{\alpha\beta})(K^{\alpha\beta} - iL^{\alpha\beta})\Psi \\
&= \tfrac{1}{2}\chi^2 (\mathbf{K}^2 + \mathbf{L}^2)\Psi \\
&= \chi^2 (\mathbf{K}^2 - \tfrac{1}{2}J^2)\Psi \\
&= \chi^2 \mathbf{K}^2 \Psi + \chi^2\{\tfrac{1}{2}I_{\alpha\beta}I^{\alpha\beta} + \tfrac{1}{2}I_{\dot\lambda\dot\mu}I^{\dot\lambda\dot\mu}\}\Psi \\
&= \chi^2\{s(s+1) + \tfrac{1}{2}I_{\alpha\beta}I^{\alpha\beta} + \tfrac{1}{2}I_{\dot\lambda\dot\mu}I^{\dot\lambda\dot\mu}\}\Psi,
\end{aligned}$$

since in the rest system $\mathbf{K}^2\Psi = s(s+1)\Psi$ if the particle has the 'pure' spin s. Therefore

$$[\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} I_{\alpha\beta} I_{\dot\lambda\dot\mu} - \chi^2(s(s+1) + \tfrac{1}{2}I_{\alpha\beta}I^{\alpha\beta} + \tfrac{1}{2}I_{\dot\lambda\dot\mu}I^{\dot\lambda\dot\mu})]\Psi = 0. \tag{137}$$

Now (137) is a spinor equation and therefore since it holds in the rest system it must hold in every system. (137) is therefore the condition which must be satisfied in order that the particle may possess the 'pure' spin s. However, there is no real reason for preferring (137) to the simpler condition

$$[\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} I^{(u)}_{\alpha\beta} I^{(u)}_{\dot\lambda\dot\mu} - \chi^2(s(s+1) + \tfrac{1}{2}I^{(u)}_{\alpha\beta}I^{\alpha\beta(u)} + \tfrac{1}{2}I^{(u)}_{\dot\lambda\dot\mu}I^{\dot\lambda\dot\mu(u)})]\Psi = 0, \tag{138}$$

where $\quad I^{(u)}_{\alpha\beta} = \tfrac{1}{2}(u_\alpha\partial_\beta + u_\beta\partial_\alpha) \quad$ and $\quad I^{(u)}_{\dot\lambda\dot\mu} = \tfrac{1}{2}(u_{\dot\lambda}\partial_{\dot\mu} + u_{\dot\mu}\partial_{\dot\lambda}).$

(138) implies that in the rest system each element of Ψ is a series all whose coefficients lying outside the subspace $\mathfrak{R}_s$ vanish. In the case (128) the elements Ψ_ρ themselves transform according to $\mathfrak{D}_{\frac{1}{2},0}$ and $\mathfrak{D}_{0,\frac{1}{2}}$ with respect to the matrix index and therefore the entire set of components of the wave function in the rest system can be reduced into two subsets transforming according to $\mathfrak{D}_{s+\frac{1}{2}}$ and $\mathfrak{D}_{s-\frac{1}{2}}$ under the rotation group. Clearly a similar argument is applicable to (133). Now (138) can be written as

$$\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} u_\alpha u_{\dot\lambda} \partial_\beta \partial_{\dot\mu}\Psi + \chi^2\{(\kappa + \kappa^* - s)(\kappa + \kappa^* + 1 + s)\}\Psi = 0.$$

Since in our case $\kappa + \kappa^* + 1 = \lambda + 1 = 0$ it follows that (138) is equivalent to

$$\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} u_\alpha u_{\dot\lambda} \partial_\beta \partial_{\dot\mu}\Psi - \chi^2 s(s+1)\Psi = 0. \tag{139a}$$

The following equations are easily obtained from (139 *a*):

$$\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} \partial_\alpha \partial_{\dot\lambda} u_\beta u_{\dot\mu}\Psi - \chi^2 s(s+1)\Psi = 0, \tag{139b}$$

$$\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} u_\alpha \partial_{\dot\lambda} \partial_\beta \partial_{\dot\mu}\Psi - \chi^2 (s-j)(s+j+1)\Psi = 0, \tag{139c}$$

$$\partial^{\alpha\dot\lambda}\partial^{\beta\dot\mu} \partial_\alpha u_{\dot\lambda} u_\beta \partial_{\dot\mu}\Psi - \chi^2 (s+j)(s-j+1)\Psi = 0. \tag{139d}$$

These four equations are completely equivalent. It is easily seen that they can be satisfied by imposing any one of the following four subsidiary conditions on Ψ:

$$\partial^{\beta\dot{\mu}}\partial_{\beta}\partial_{\dot{\mu}}\Psi = 0 \quad (s = j = 0), \tag{140a}$$

$$\partial^{\beta\dot{\mu}}u_{\beta}u_{\dot{\mu}}\Psi = 0 \quad (s = j = 0), \tag{140b}$$

$$\partial^{\beta\dot{\mu}}\partial_{\beta}u_{\dot{\mu}}\Psi = 0 \quad (s = j \geqslant 0), \tag{140c}$$

$$\partial^{\beta\dot{\mu}}u_{\beta}\partial_{\dot{\mu}}\Psi = 0 \quad (s = -j \geqslant 0). \tag{140d}$$

(140*a*) and (140*b*) are equivalent since one goes over into the other by the substitution $\kappa \rightarrow -\kappa - 1, \kappa^* \rightarrow -\kappa^* - 1$. The same is true of (140*c*) and (140*d*). The equation proposed by Dirac (1945) for a spinless particle corresponds to (133) and (140*b*), using the five-row representation for β_k.

Notice that (137) or (138) ensure that Ψ_ρ reduces to a finite series in the rest system. Therefore as proved in § 5 it is continuously differentiable an arbitrary number of times. In an arbitrary system Ψ_ρ is in general an infinite series having components in every subspace $\mathfrak{R}_k$. This, however, in my opinion, does not justify the interpretation given by Dirac (1945), that the particle develops new spins during motion. Even in the case of finite representations Ψ has components lying outside the subspace $\mathfrak{R}_s$ in an arbitrary frame of reference. This is usually not interpreted to mean that the spin of the particle is altered by motion.

Finally, I should like to thank Professor Dirac for suggesting the investigation of these infinite representations.

REFERENCES

Courant & Hilbert 1931 *Methoden der Mathematischen Physik*, **1**, chapter II. Berlin: Springer.
Dirac 1936 *Proc. Roy. Soc.* A, **155**, 447–459.
Dirac 1945 *Proc. Roy. Soc.* A, **183**, 284–295.
Fierz 1939 *Helv. Phys. Acta*, **12**, 3–37.
Harish-Chandra 1947 *Proc. Camb. Phil. Soc.* (in the Press).
Pauli 1940 *Phys. Rev.* 58, 716–722.
Pontrjagin 1939 *Topological groups*. Princeton: Princeton University Press.
Van der Waerden 1932 *Die gruppentheoretische methode der Quantummechanik*. Berlin: Springer.
Weyl 1931 *The theory of groups and quantum mechanics*. London: Methuen.
Weyl 1939 *The classical groups*. Princeton: Princeton University Press.

Thermal diffusion in mixtures of the inert gases

By K. E. Grew, *Heriot-Watt College, Edinburgh*

(*Communicated by M. L. Oliphant, F.R.S.—Received* 15 *August* 1946)

Measurements have been made of the separation produced by thermal diffusion in all the binary mixtures of the inert gases (except Kr-X and Rn mixtures) at temperatures ranging from −180 to 400° C. The temperature dependence of the thermal diffusion ratio, k_T, and of $R_T (R_T = k_T/[k_T(\infty)]_1$, where $[k_T(\infty)]_1$ is the theoretical value for molecules behaving like rigid elastic spheres) has been deduced. The results indicate that, in general, R_T increases with the temperature in the lower range and becomes constant at high temperatures. The constant value is first attained at a temperature which varies from mixture to mixture, but its magnitude is nearly the same for all mixtures. It corresponds with a repulsive force index of 11, if the molecules are treated as centres of repulsive force only.

The results have been compared with those calculated by R. Clark Jones (1940, 1941) for a gas mixture in which the two molecular species have identical force fields of the Lennard-Jones type, with repulsive force index $\nu = 9$ and attractive force index $\nu' = 5$. The theoretical results give R_T as a function of T/T_C, where T_C is the critical temperature of the gas. This relation with the critical temperature is supported by the experimental results, though there are difficulties in the comparison due to the non-identical nature of the force fields in the experimental mixtures. It appears, however, that the observed behaviour could be explained by a Lennard-Jones model if the force indices were suitably chosen, and that a study of the temperature dependence of R_T in isotopic mixtures would enable the indices to be determined with some precision.

INTRODUCTION

The work to be described is an experimental investigation of the dependence on temperature of the thermal diffusion ratio k_T for the various combinations of the inert gases, and a comparison of the results with the theoretical ones of R. Clark Jones (1940, 1941) for molecules which have an attractive field superimposed on a repulsive field.

The thermal diffusion ratio k_T appears in the Chapman-Enskog equation relating the concentrations c_1, c_2 of the molecular species in a binary gas mixture in which a temperature gradient exists to the temperature

$$\nabla c_1 = -\nabla c_2 = -k_T . \nabla \log T.$$

k_T is a complicated function of the masses, concentrations and force fields of the molecules. The effect of the fields enters in the form of integrals Ω which may be considered as dimensionless collision cross-sections, and unless the two molecular species have identical fields of force, distinction must be made between the interactions of like and unlike molecules, that is, between Ω_1, Ω_2 and Ω_{12}. When the molecules have identical fields, as in mixtures of isotopes, only one type of interaction occurs. Chapman describes such molecules as 'isotopic', whether or not they are isotopes of the same element. Prior to Clark Jones's work k_T had been evaluated only in the following special cases:

(i) for molecules interacting like rigid elastic spheres of different diameters σ_1, σ_2;

(ii) for mixtures of isotopic molecules which are point centres of repulsive force F varying inversely as the νth power of the distance r, i.e. $F = \kappa r^{-\nu}$.

In both these cases k_T is independent of the temperature.

Experimental determinations of k_T have been made for various gas mixtures (Ibbs & Grew 1931; Atkins, Bastick & Ibbs 1939, etc.; Grew 1941). A comparison of the experimental and theoretical values enables an estimate to be made of the force index ν, assuming the interaction to be an inverse power repulsion. In many cases, however, the value of k_T changes with temperature. This is an indication that an inverse power repulsion does not adequately represent the molecular fields, since for this type of interaction k_T is theoretically independent of the temperature if the molecules are isotopic, and only slightly dependent otherwise (Chapman 1940).

As a result of recent experimental work on thermal diffusion in mixtures of isotopes (Nier 1940), in which also a dependence of k_T on temperature was observed, Clark Jones (1940, 1941) has treated theoretically thermal diffusion in isotopic mixtures in which the interaction includes an inverse power attraction superimposed on an inverse power repulsion, thus $F = \kappa r^{-\nu} - \kappa' r^{-\nu'}$ (Lennard-Jones's model). This model is on first consideration unpromising as a means of explaining the observed variation of k_T with temperature, which is a decrease with decreasing temperature: it suggests, indeed, that the variation should be in the opposite direction. For, considering a repulsive field only, $F = \kappa r^{-\nu}$, it is seen that

$$\nu = -d \log F / d \log r.$$

This quantity Clark Jones calls the hardness function $\mu(\nu)$. For the Lennard-Jones model it is

$$\mu(\nu) = \nu + Q(\nu - \nu') + O(Q^2), \quad \text{where} \quad Q = \frac{\kappa'}{\kappa} r^{(\nu-\nu')}.$$

For values of r less than r_0 (the value of r at which the potential energy is a minimum), Q is positive and less than unity, so that $\mu(\nu) > \nu$, and by an amount which decreases as r decreases, that is, as the temperature increases. It would be expected then that k_T would decrease with increasing temperature, contrary to what is observed. Clark Jones has shown that this is the case only at high temperatures; at lower temperatures the effect of the attractive field is to cause k_T to decrease as the temperature falls, and even to change sign.

His fullest treatment is for an interaction in which the repulsive index $\nu = 9$, and the attractive index $\nu' = 5$, a model which was chosen because much of the numerical work had already been done. His results are shown graphically in figure 1. Here the ordinate R_T is the ratio of k_T for the particular model to the value for rigid elastic spheres $k_T(\infty)$. (R_T is more conveniently discussed than k_T, since it is roughly independent of the concentrations of the molecular species and only slightly dependent on their masses, so that its value is determined mainly by the molecular fields.) The abscissa is the quantity ϵ/kT, ϵ being the depth of the potential energy minimum, i.e.

$$\epsilon = \frac{\nu - \nu'}{(\nu - 1)(\nu' - 1)} \frac{\kappa'^{(\nu-1)/(\nu-\nu')}}{\kappa^{(\nu'-1)/(\nu-\nu')}},$$

and k is the Boltzmann constant. Clark Jones has compared these theoretical values with some experimental values of R_T obtained by Nier (1940) for mixtures of the isotopes ^{20}Ne-^{22}Ne, and $^{12}CH_4$-$^{13}CH_4$. The observed values are mean values over a wide temperature range, and the maximum value is considerably greater than the maximum theoretical value, presumably because the actual repulsive index is greater than 9. Despite these difficulties the comparison suggests that the observed variation of R_T could reasonably be explained with the Lennard-Jones model, though the indices (9, 5) are not suitable, at least for neon.

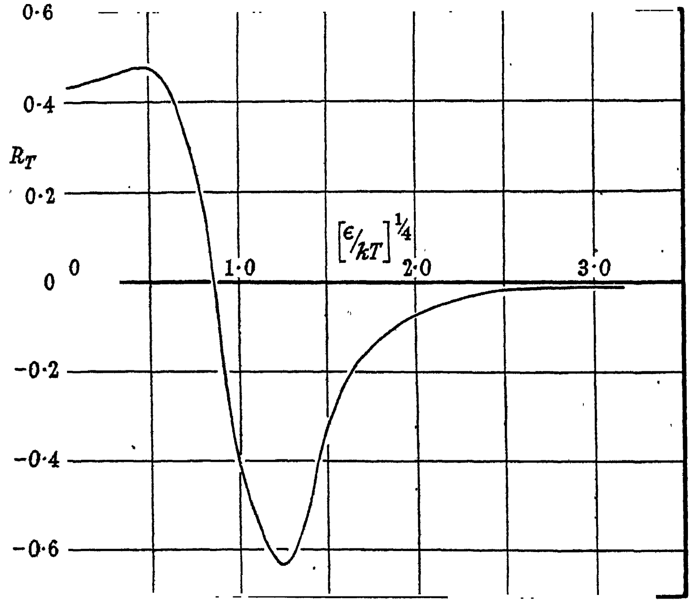


FIGURE 1. The variation of R_T with temperature for the (9, 5) model (Clark Jones). The abscissa is approximately $[T_C/T]^{\frac{1}{4}}$, where T_C is the critical temperature of the gas.

As a result of this extension of the theory of the thermal diffusion effect further measurements of the variation of R_T with temperature are desirable. These should preferably apply to isotopic mixtures, since in them only one type of interaction has to be considered. But, although for such mixtures the theoretical treatment is relatively simple, the experimental difficulties are large, and the work which is now described is a measurement of the thermal diffusion effect over a wide range of temperature in mixtures of the inert gases—which presumably form the next simplest case to that of isotopic mixtures.

Some measurements on mixtures of the inert gases have already been reported. Atkins *et al.* (1939) have examined all the binary combinations of He, Ne, A, Kr and X, but the effect was measured only for a temperature gradient from 15 to 100° C. Measurements over a range of temperature from 0 to −180° C were made by Ibbs & Grew (1931) for He-Ne, He-A and Ne-A mixtures. These indicated clearly that R_T diminishes as the temperature falls. Since this work was done the inert gases have come to be commercially obtainable in a pure state, and larger volumes may now be used than formerly. A modification of the original method has been made to permit measurements on one and the same mixture over a temperature range from − 180 to 400° C.

EXPERIMENTAL METHOD

In general, the method and procedure were the same as in the study of the effect in hydrogen-deuterium mixtures (Grew 1941). The analysis of the gas mixture before and after separation was made by a small thermal conductivity gauge to which a sample of gas of volume 0·2 to 0·4 c.c. was admitted from the diffusion apparatus by means of a special sampling cock. The diffusion vessel consisted of a large bulb *A*, of volume 61·3 c.c., connected by a tube 9·5 cm. long and 3·6 mm. internal diameter to a small bulb *B* which formed part of a stopcock, as shown in figure 2. The volume of *B* was 2·6 c.c. Diffusion proceeded with the stopcock turned so that the short side tube *C* was in communication with the bulbs. Rotation of the cock then spilt the gas in *C* into the gauge for analysis.

The gas mixtures were prepared in known proportions in a burette attached to the apparatus and were then admitted to the evacuated bulbs. The pressure was adjusted to about 30 cm. of mercury at room temperature prior to the high-temperature measurements, and to about 70 cm. prior to the low-temperature measurements. This ensured that at high temperatures the pressure did not exceed atmospheric pressure, and at low temperatures there was sufficient gas in the side tube *C* for analysis. The separation was thus measured at pressures which varied with the temperature, but theoretically the separation is independent of the pressure and experiment has confirmed this (Grew 1945).

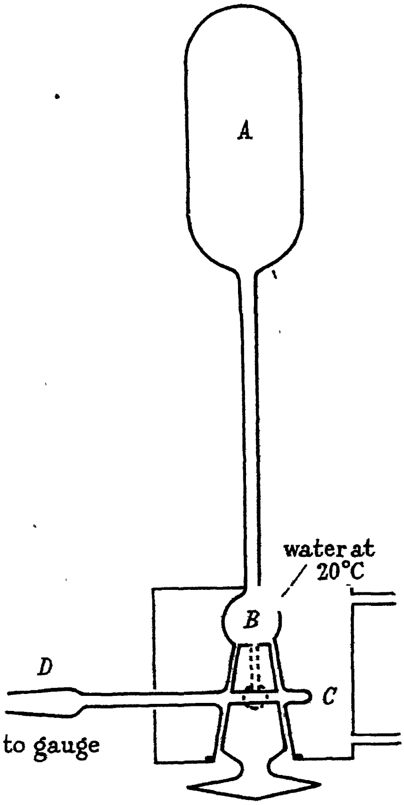


FIGURE 2. The diffusion bulbs.

For high temperatures the bulb *A* was placed in an electric furnace. The support for the nichrome winding was a thick-walled copper tube on which there were supplementary windings at either end so that the region near the bulb could be brought to a uniform temperature. The low-temperature measurements were made in two ways. In the first method the bulb *A* (which was now brought below *B* by rotation about the ground-joint *D*) was surrounded by a copper cylinder, a little longer than the bulb, on which a few turns of nichrome wire were wound to form a heating coil. The cylinder was supported inside a double-walled vessel of which the space between the walls could be evacuated. Surrounding this vessel was a vacuum vessel containing liquid oxygen. Temperatures from -120 to $-180°$ C were thus readily maintained by adjustment of the heating current and of the vacuum between the walls of the inner vessel. In the second method, applicable between -20 and $-150°$ C, the bulb was surrounded with pentane contained in a vacuum vessel, and cooled and maintained at constant temperature by addition of liquid oxygen or solid carbon dioxide. Manual control sufficed to keep the temperature constant to within 0·5° C for the 45 min. or so required for equilibrium to be attained and the samples analyzed. Over the range of temperature common to both methods excellent agreement was obtained. In those cases in which the boiling-point of the gases permitted of it, the bulb was immersed in liquid oxygen for the lowest temperature.

Temperatures were measured by three copper-constantan couples placed close to the upper, middle and lower parts of the bulb A respectively, so that the temperature variation over the bulb could be measured and a mean temperature found. Above 300° C the variation was 4° C; at low temperatures not more than 2° C. The couple was calibrated frequently, using as reference points the boiling-point of oxygen, the freezing-point of mercury, and the boiling-points of water, naphthalene and benzophenone. The temperature of the water circulating about the small bulb was measured with a mercury thermometer. The temperature was controlled to within 0·5° C.

Measurement of the separation. The difference in composition of the gas in the large bulb and in the small bulb when a temperature gradient exists in the connecting tube is termed the separation. Only the change in composition in the small bulb was measured, the separation being calculated from this and the relative volumes and temperatures of the two bulbs. Both bulbs were first brought to the same temperature and a sample of gas admitted to the gauge, which, as before (Grew 1941), formed one arm of a bridge across which the potential difference could be varied. The potential difference to balance the bridge for the uniform mixture was thus found. The large bulb was then brought to the required temperature, and after 20 min. or more a sample was withdrawn from the small bulb. From the change in potential difference required to balance the bridge and the slope of the potential difference-concentration curve, the change in composition of the gas in the small bulb was calculated. To ensure that equilibrium had been attained a second sample was taken later. The separation is then given by

$$s = s'\left(1+\frac{V_1 T_2}{V_2 T_1}+\theta\right),$$

where s' = change in composition in the small bulb; V_1 = volume of small bulb; V_2 = volume of large bulb; T_1 = temperature of small bulb; T_2 = temperature of large bulb, and θ = a correction term for the gas in the connecting tube. θ is a function of the volumes of the connecting tube and large bulb and of the temperatures T_1 and

T_2. In finding θ it was assumed that the temperature gradient in the connecting tube was uniform and that k_T could be regarded as constant throughout the gas in the tube. Both assumptions were sufficiently valid for the calculation of the correction. θ increases with the temperature T_2; at the highest temperatures used its value was 0·013.

It is estimated that in mixtures containing He or Ne, the error of a measurement of the percentage separation is ± 0·03. The gauge is less sensitive for A-Kr and A-X mixtures, and here the error may be ± 0·05.

Preparation of the gases. The inert gases used were all supplied by the British Oxygen Co. The helium, neon and argon were spectroscopically pure. The krypton was of German origin and stated to contain 1 % xenon. No attempt was made to purify this gas. The xenon was produced in the U.S.A., and, as its purity was not given, the gas was fractionated. The gauge readings for the various fractions indicated that the gas contained a small amount of impurity—about 2 % if it was krypton and a smaller amount if it was a lighter gas. The residue, which alone was used in the diffusion measurements, contained without doubt only a negligible amount of impurity.

RESULTS

The values of the separation, when plotted against $\log T_2/T_1$ (T_1 = the constant temperature of the small bulb, T_2 = the varying temperature of the large bulb), yield in some cases and over a limited range of temperature T_2 a straight line, indicating that the fundamental equation

$$dc = k_T d(\log T)$$

may be integrated as if k_T were constant, to give the separation

$$s = c_2 - c_1 = k_T \log T_2/T_1$$

(where c_1, c_2 are the concentrations at temperatures T_1, T_2). In general, however, the linear relation between the separation and $\log T_2/T_1$ does not hold if T_2 is sufficiently reduced, and it is necessary to consider the interpretation of the separation $-\log T_2/T_1$ curve in this case.

k_T is a function of the concentration, and for some types of molecular field of the temperature also. If the separation is not too large the dependence on concentration may be ignored and k_T regarded as a function of temperature only. Then

$$s = c_2 - c_1 = \int_{T_1}^{T_2} k_T d(\log T) \quad \text{and} \quad ds = \int_{T_2}^{T_2+dT_2} k_T d(\log T),$$

since T_1 is constant in this method of experiment. Hence

$$ds = k_T d(\log T_2).$$

The slope of the separation $-\log T_2/T_1$ curve at any point is therefore a measure of k_T at the temperature T_2 corresponding to that point.

The assumption made here that k_T may be considered as independent of the concentration in the range of separation involved is questionable, as the separation amounts in some of the mixtures examined to 10 %. It is, however, not easily avoided, and its effect was therefore minimized. First, the mixtures used were those containing nearly equal proportions of the two components and for which therefore the variation of k_T with composition is small. Secondly, each observed separation was corrected so that it applied to a mixture of mean composition the same as that of the original uniform mixture. In the form of apparatus used, most of the change in composition occurs in the small bulb. This causes the mean composition of the gas in the connecting tube to vary with the separation attained, and consideration shows that there is an apparent decrease in $k_T = ds/d(\log T_2)$ as T_2 increases, since in all cases the proportion of the lighter component was less than that for maximum value of k_T. The reverse effect would occur if the change in composition occurred in the large bulb. For equal changes in composition in the two bulbs, that is, for constant mean composition of the gas in the connecting tube, the separation $-\log T_2/T_1$ curve should be linear. Each value of the separation was therefore corrected so that it applied to a mixture of mean composition equal to that of the original mixture. For this purpose use was made of the curves showing the separation as a function of the composition (for constant temperature ratio) given by Atkins *et al.* (1939). It is these corrected values which are recorded and from which the values of k_T were derived. That this treatment is satisfactory is shown by the fact that for those mixtures in which k_T is expected to be independent of temperature (for example, He-Ne except at very low temperatures) the separation $-\log T_2/T_1$ plot is a straight line even though the separation amounts to 8 % or more. The apparent variation of k_T in other cases is therefore a temperature effect.

The results are summarized in table 1, where the values of the separation at definite values of $\log T_2/T_1$ are given for the nine mixtures examined.* R_T is the ratio of k_T (determined from the slope of the separation $-\log T_2/T_1$ curves) to $k_T(\infty)$, this being the theoretical value of k_T for the same mixture assuming that the molecules interact like rigid elastic spheres of diameters equal to those deduced from the viscosity of the pure gases. Expressions for $k_T(\infty)$ have been given by Atkins *et al.* (1939). The value of R_T given in the last column is that found by Atkins *et al.* for a mixture of the same composition; it is a mean value in the range 15 to 100° C. There is reasonably good agreement except in the case of He-Ne. The difference between the present value for this mixture of 0·64 and Atkins *et al.* value of 0·80 is much greater than the experimental error. The cause of this discrepancy is unknown. In case the gases used had been inadvertently contaminated during manipulation, new supplies were obtained and further measurements made. The following are the results:

mixture	% He	k_T	$k_T(\infty)$	R_T
1	36·8	0·070	0·104	0·67
2	53·8	0·080	0·124	0·64
3	62·5	0·078	0·125	0·62

---

* The method of analysis was not sufficiently sensitive to measure reliably the small separation in Kr-X mixtures.

TABLE 1

mixture	$[k_T(\infty)]_1$	$\log_{10} T_2/T_1$	−0·5	−0·4	−0·3	−0·2	−0·1	0·0	0·1	0·2	0·3	ABI
		t_2 °C	−180	−156	−126	−88	−40	20	96	191	312	
He-Ne	0·124	sepn.	−8·65	−7·02	−5·30	−3·60	−1·80	0·0	1·80	3·60	5·42	
53·8 % He		R_T		0·58	0·60	0·63	0·64	0·64	0·64	0·64	0·64	0·80
He-A	0·151	sepn.	−10·20	−8·35	−6·35	−4·40	−2·18	0·0	2·18	4·40	6·70	
51·2 % He		R_T		0·55	0·56	0·61	0·63	0·64	0·65	0·66	0·66	0·63
He-Kr	0·166	sepn.		−9·70	−7·50	−5·10	−2·60	0·0	2·52	5·10	7·64	
55·0 % He		R_T		0·55	0·61	0·64	0·67	0·67	0·67	0·67	0·67	0·63
He-X	0·163	sepn.				−4·90	−2·50	0·0	2·52	5·02	7·53	
53·6 % He		R_T				0·66	0·66	0·66	0·66	0·66	0·66	0·62
Ne-A	0·084	sepn.	−3·68	−3·21	−2·63	−1·83	−0·95	0·0	1·09	2·20	3·30	
51·2 % Ne		R_T		0·28	0·36	0·44	0·48	0·52	0·57	0·57	0·57	0·55
Ne-Kr	0·132	sepn.		−5·00	−4·02	−2·93	−1·62	0·0	1·75	3·58	5·47	
53·0 % Ne		R_T		0·30	0·34	0·40	0·47	0·55	0·59	0·61	0·64	0·53
Ne-X	0·152	sepn.				−3·30	−1·72	0·0	1·85	3·88	6·15	
54·2 % Ne		R_T				0·43	0·46	0·49	0·54	0·60	0·66	0·44
A-Kr	0·079	sepn.		−0·95	−0·84	−0·65	−0·38	0·0	0·52	1·23	2·20	
53·5 % A		R_T			0·09	0·12	0·18	0·24	0·33	0·47	0·59	0·19
A-X	0·116	sepn.				−0·83	−0·47	0·0	0·62	1·50	2·60	
56·4 % A		R_T				0·13	0·15	0·18	0·29	0·37	0·45	0·17

The apparent dependence of R_T on the composition is, at least in its direction, in accord with theoretical predictions for molecules exerting an inverse power repulsion (Chapman 1940, p. 58). The higher value of 0·80 is supported by Blüh, Blüh & Puschner (1937); on the other hand, values of R_T deduced from the viscosity of the pure gases (Clark Jones 1940, p. 116) favour the lower value found in the present work.

DISCUSSION OF RESULTS

The results confirm the earlier ones (Ibbs & Grew 1931) in that they show that in general R_T decreases as the temperature falls. In some mixtures, such as He-Ne, R_T is constant over almost the whole temperature range; in A-Kr R_T shows a continuous variation over the whole range. Other mixtures show a more or less extended temperature range in which R_T is constant, and below this a range in which R_T diminishes as the temperature falls. It is a reasonable conclusion that this behaviour would be shown by all the mixtures were the temperature range sufficiently extended. The constancy of R_T over a range of temperature indicates that the molecules are, in that range, behaving as if they exerted repulsive forces only.

The maximum value of R_T and the corresponding force index. It is remarkable that, although the temperature at which R_T first becomes constant varies widely from mixture to mixture, the value of R_T when constant, that is, the maximum value—R_T (max.), is much the same for all mixtures, viz. about 0·64. Thus the change in R_T in passing from one to another of the inert gas mixtures, observed by Atkins *et al.* (1939), is now seen to be an accident of the temperature range to which their measurements were restricted. It appears that at sufficiently high temperatures R_T would be nearly the same for all.

Assuming that, at high temperatures, the molecules behave like centres of repulsive force only, $F = \kappa r^{-\nu}$, an estimate can be made of the force index ν by comparing the experimental value of R_T (max.) with the theoretical values of R_T for different values of ν. A difficulty arises in the comparison, however, in that theoretical values of R_T are known exactly only for the special case of a mixture in which the ratio of the molecular masses is very large and the concentration of the heavier molecule tends to unity (the Lorentzian case).* In the general case only the ratio of the first approximations to k_T, $[k_T(\nu)]_1/[k_T(\infty)]_1$, is known, whereas the theoretical value of R_T is more properly given by $k_T(\nu)/[k_T(\infty)]_1$ (where $k_T(\nu)$ is the exact value of k_T) to correspond with the experimental value of R_T defined as

$$k_T\,(\text{experimental})/[k_T(\infty)]_1.$$

For small values of the mass ratio and of ν the error of the first approximation is small, but it increases to about 23 % in the Lorentzian case with $\nu = \infty$. In the absence of an exact knowledge of the variation of the error, it may be assumed that it increases linearly with the proportionate mass difference defined by

---

* Values of ν have sometimes been estimated from theoretical values of R_T for this special case, but more recent work (Grew 1941) has shown that this leads to considerable error, and the procedure here adopted is preferable.

$$M = (m_1 - m_2)/(m_1 + m_2).$$

The error has been estimated for the case $M = 0$ and for $\nu = 5, 9, 13, 17, \infty$ (Grew 1941, p. 397), and it is known in the Lorentzian case ($M = 1$) for these same values of ν. Linear interpolation thus gives the error for intermediate values of M and ν. Then using Chapman's (1940, p. 58) data, curves can be drawn showing R_T as a function of ν for various values of M. From these curves the values of ν shown in table 2 were found to correspond with the experimental values of R_T. As the mixtures are not isotopic, ν derived in this way must be regarded as a mean of ν_1, ν_2, ν_{12}. The possible error of 0·01 in R_T introduces an uncertainty of 0·4 in the values of ν.

TABLE 2. VALUES OF THE PROPORTIONATE MASS RATIO M, OF R_T AND OF THE CORRESPONDING FORCE INDEX ν

mixture	He-Ne	He-A	He-Kr	He-X	Ne-A	Ne-Kr	Ne-X	A-Kr	A-X
M	0·67	0·82	0·91	0·94	0·33	0·61	0·73	0·36	0·53
R_T(max.)	0·64	0·66	0·67	0·66	0·57	0·64			
ν	11·0	10·9	10·8	10·5	10·9	11·2			

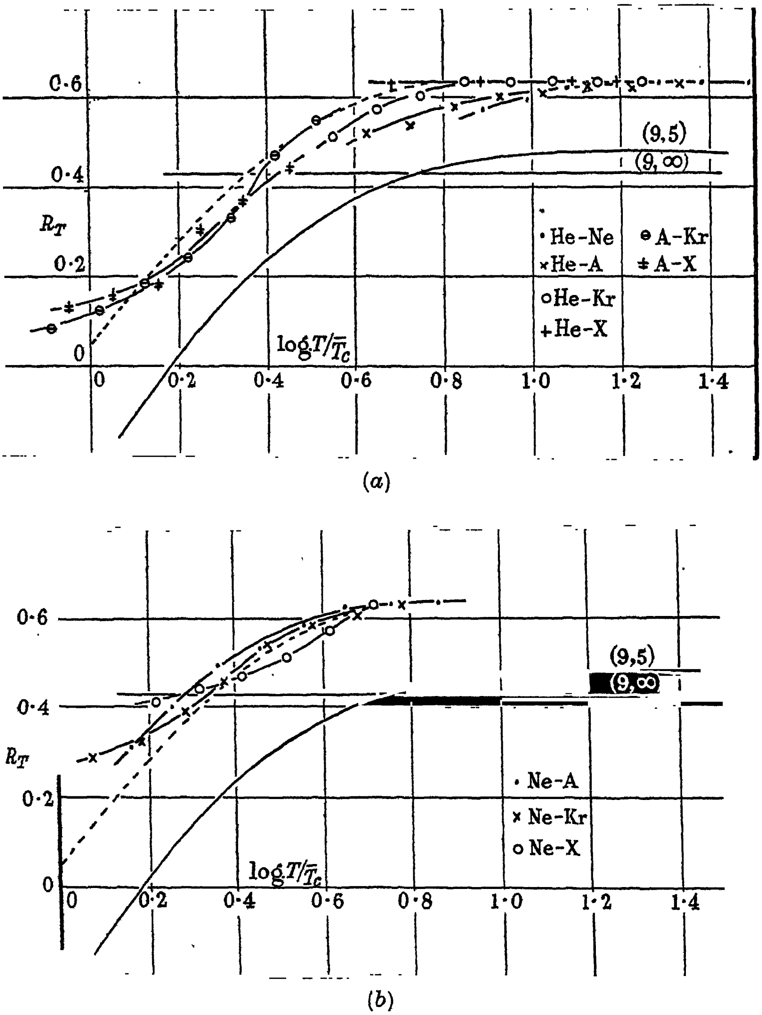
Thus in the six mixtures for which values can be stated ν is close to 11. This suggests that, in so far as they can be represented by an inverse power repulsion, that is, at high temperatures, the fields of the inert gas atoms are alike in having the same, or nearly the same, force index and differ only in the force constant.

The variation of R_T with temperature. Reference has already been made to Clark Jones's (1940, 1941) calculation of R_T for the Lennard-Jones molecular field, which includes an inverse power attraction superimposed on a repulsive force: thus $F = \kappa r^{-\nu} - \kappa' r^{-\nu'}$. The calculated values refer to the special case of $\nu = 9$, $\nu' = 5$. R_T is given as a function of a quantity ϵ/kT, where k is the Boltzmann constant, T the absolute temperature, and ϵ is the minimum potential energy of two interacting molecules. According to Lennard-Jones and Devonshire, ϵ is related to the critical temperature T_C of the gas—$\epsilon = kT_C$, or better $\epsilon = kT_C/1{\cdot}22$ (cf. Fowler & Guggenheim 1939, p. 345). Using this latter relation, the curve marked (9, 5) of figure 3*a*, *b* has been drawn, representing R_T as a function of $\log T/T_C$. The curve $(9, \infty)$ is that for a molecule which exerts a repulsive force only, with index 9. In both cases the values of R_T apply to an isotopic mixture in which M is small. Also, R_T is here the ratio of the first approximations $[k_T(9,5)]_1/[k_T(\infty)]_1$.

To compare the experimental results with these theoretical ones some assumption must be made about the 'critical temperature' to be assigned to the gas mixtures since they are not isotopic. Clark Jones gives reasons for taking as the critical temperature $\bar{T}_C$ of a mixture the mean proportional of the critical temperatures of the two components. If this is done the following values are found:

mixture	He-Ne	He-A	He-Kr	He-X	Ne-A	Ne-Kr	Ne-X	A-Kr	A-X
$\bar{T}_C$ °K	15	27·5	32·5	38	81·5	96	112	178	209

The experimental values of R_T at temperature T ($T = T_2$ of table 1) may now be plotted against $\log T/\overline{T}_C$. But as R_T (max.) is not the same for all the mixtures, partly because of the dependence on the mass ratio, the maximum values of R_T for each mixture (except A-Kr, A-X, for which a constant value of R_T is not reached) have first been made equal to 0·63 and the other values changed in proportion. These are the values shown in figure 3*a*, *b*.



(*a*)

(*b*)

FIGURE 3*a*, *b*. R_T as a function of $\log T/\overline{T}_C$. Experimental values ⊙ × etc. (9, 5): Clark Jones's theoretical values. (9, ∞): for repulsive field only, index 9. Broken curve: a reference curve for comparison of figures 3*a*, *b* and 4.

It is apparent from the curves that:

(1) The experimental values of R_T are higher than the theoretical value for the same value of $\log T/\overline{T}_C$. In particular, the maximum experimental value is greater

than the maximum theoretical value, and since in this temperature range the attractive field has little effect, the discrepancy is an indication that the repulsive force index in these mixtures is greater than 9. It has, indeed, already been estimated to be 11 when a repulsive field only is assumed to be operative. This value would be somewhat smaller for a Lennard-Jones field, for Clark Jones's results show that in this temperature range the effect of the attractive field is to increase R_T above the value for the repulsive field only.

(2) In some cases, such as Ne-A, the form of the experimental curve is strikingly similar to that of the theoretical curve. More frequently, however, R_T changes less rapidly at the lowest temperatures than at the intermediate ones, so that the experimental curve becomes convex to the temperature axis. The explanation of this behaviour may perhaps be given when there are available theoretical data for force indices other than (9, 5). It is a reasonable supposition, however, that as ν' increases above 5, R_T changes sign, if it does so at all, at progressively lower temperatures. As a probable value of ν' is 7, this would explain the fact that the experimental results give no indication of R_T becoming negative.

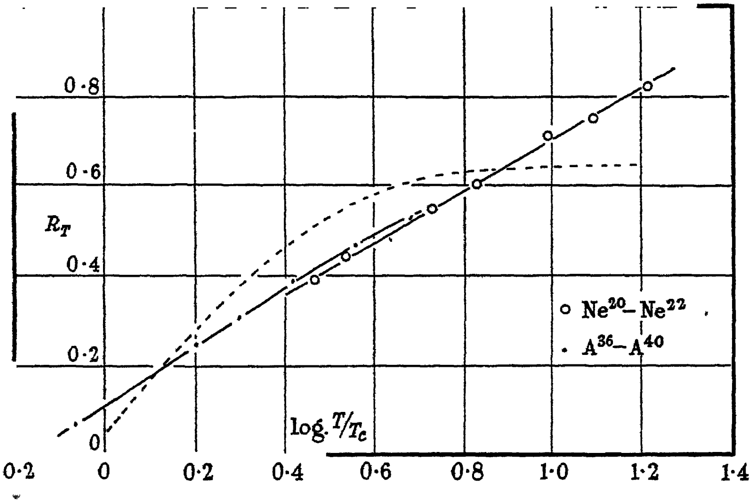


FIGURE 4. Stier's experimental values for Ne isotopes and for A isotopes. Broken curve: as in figure 3.

(3) The experimental curves are coincident (except at the lowest temperatures) to a degree which is significant when the range of the 'critical temperature' and the method by which it was calculated are taken into consideration. The formal relation between R_T and the critical temperature brought out by Clark Jones has therefore this much experimental support.

The variation of R_T in isotopic mixtures. It is interesting to compare with these results for non-isotopic mixtures the measurements of R_T in the isotopic mixtures Ne^{20}-Ne^{22} and A^{36}-A^{40} made by Stier (1942). The curves of figure 4 show Stier's values of R_T as a function of T/T_C. The A values, and the Ne values too in the low-temperature range, conform satisfactorily with the others, when it is remembered that the mass ratio M in these cases is only 0·05 and R_T may in consequence be 10 to 20 % smaller than for the non-isotopic mixtures with their larger mass ratio. The persistent increase of R_T at the higher temperatures in the case of the Ne isotopes is unexpected.

In spite, then, of the difficulty introduced into their interpretation by the non-isotopic nature of the mixtures, it seems that the experimental results could be satisfactorily explained with molecular fields of the Lennard-Jones type if the force indices were suitably chosen. The results suggest, indeed, that the measurement of the temperature variation of R_T in isotopic mixtures would permit of the determination of both indices, ν, ν', with some precision. A knowledge of these values combined with the treatment of the second virial coefficient due to Lennard-Jones (1931) and Buckingham (1938) would enable the force constants κ, κ' to be determined, and the molecular fields would then be completely known. Further progress depends on the provision of more experimental data for isotopic mixtures, and of theoretical data for molecular fields with values of (ν, ν') other than $(9, 5)$.

I acknowledge with gratitude a grant from the Carnegie Trust for the Universities of Scotland for the purchase of apparatus and materials. I am indebted also to Dr R. Fürth and to Dr T. L. Ibbs for discussions and criticism.

REFERENCES

Atkins, B. E., Bastick, R. E. & Ibbs, T. L. 1939 *Proc. Roy. Soc.* A, **172**, 142.
Blüh, G., Blüh, O. & Puschner, M. 1937 *Phil. Mag.* **24**, 1103.
Buckingham, R. A. 1938 *Proc. Roy. Soc.* A, **168**, 264.
Chapman, S. 1940 *Proc. Roy. Soc.* A, **177**, 38.
Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical Thermodynamics.* Camb. Univ. Press.
Grew, K. E. 1941 *Proc. Roy. Soc.* A, **178**, 390.
Grew, K. E. 1945 *Nature*, **156**, 267.
Ibbs, T. L. & Grew, K. E. 1931 *Proc. Phys. Soc.* **43**, 142.
Jones, R. Clark. 1940 *Phys. Rev.* **58**, 111.
Jones, R. Clark 1941 *Phys. Rev.* **59**, 1019.
Lennard-Jones, J. E. 1931 *Proc. Phys. Soc.* **43**, 461.
Nier, A. O. 1940 *Phys. Rev.* **57**, 338.
Stier, L. G. 1942 *Phys. Rev.* **62**, 548.

A theory of the dependence of the rate of detonation of solid explosives on the diameter of the charge

BY H. JONES, *Imperial College, London*

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 16 *September* 1946)

The velocity of a detonation wave, in a cylindrical charge of solid explosive, is shown to be dependent on the diameter of the charge, and the relation between the velocity and the diameter is calculated. It is shown that this effect depends upon the rate of the chemical reaction occurring in the front portions of the detonation wave, and that it is possible, therefore, to determine this rate of reaction by measuring the velocity of detonation in bare charges of different diameters. The effect of a metal case surrounding the charge is also briefly discussed.

INTRODUCTION

A successful theory of the detonation of gases contained in tubes was first given by D. L. Chapman (1899) and later was developed in greater detail by Jouguet (1906). These authors were able to predict with considerable accuracy the velocity of the detonation wave in an explosive gas in terms of the heat of the reaction, and such quantities as the specific heats of the resultant gases. An interesting feature of the theory is that, not only is the velocity of the detonation wave determined, but also the pressure and the temperature immediately behind the detonation front. In accordance with many observations, the theory is based on the assumption that the detonation process is a steady one, i.e. that the velocity of the detonation wave remains constant. It is on this account that, in this theory, the rate of the reaction occurring in the initial portion of the detonation wave does not enter into the calculation of the velocity of propagation.

It is natural that the Chapman-Jouguet theory should have been applied to the detonation of cylindrical charges of solid and liquid high explosives. Many authors have attempted this extension of the theory but with only limited success. There are two main reasons why a theory of detonation in solids is more difficult than in gases. First, because the products of the explosion are, in general, more complicated; many different molecular species being formed, and, secondly, because the pressure, temperature and molecular volume in the detonation wave have values which lie far outside the range which has been studied in other connexions. Thus very little guidance is available for setting up an equation of state applicable to the conditions of the detonation wave.

Apart, however, from the difficulty associated with the equation of state, the dynamics of the Chapman-Jouguet theory require modification when applied to solid explosives. This arises from the fact that, while the detonation of a gas in a rigid tube leads to a flow which may be regarded as strictly one-dimensional, in solid high explosives no practicable confinement can prevent a certain radial expansion of the decomposing materials. As far as the determination of the velocity of detonation is concerned, only the flow within that region of the wave where the decomposition of the explosive takes place is of significance. The subsequent flow of the reaction products has no effect on the steady velocity of the detonation wave.

The effect of radial expansion within the reaction zone can easily be understood in general terms. It is known from experience that the detonation velocity of a solid explosive generally increases with the loading density. Hence, if as a result of radial expansion the density is reduced before the reaction is complete, part of the explosive decomposes at a lower effective loading density, and a velocity of detonation may be expected less than that which would occur if no radial expansion took place, and therefore less than that given by a straightforward application of the Chapman-Jouguet theory. It is with this problem of the reduction of the detonation velocity by radial expansion within the reaction zone that this paper is concerned. The main interest in this matter lies in the fact that it provides a means of determining experimentally the rate of the reaction in a detonation wave. Clearly the degree of effective radial expansion depends upon the confinement of the charge and upon the length of the reaction zone. A knowledge of the length of the reaction zone gives at once the time required for the complete decomposition of the explosive and the release of the chemical energy. In the second and third sections of this paper formulae are given from which the time of reaction can be obtained from data giving the variation of the velocity of detonation with the diameter of the charge, or, for cased charges, with the weight of the casing.

It is often convenient to consider the detonation process as judged from axes moving with the detonation wave. The front portion of this wave is simply a powerful shock wave travelling through the unexploded material. It is the compression and the temperature in this shock-wave front which initiates the reaction. If we imagine the wave to be travelling from right to left, the shock-wave front, from the moving axes, appears to be stationary and unexploded material appears to be passing through it from left to right with the detonation velocity. It is assumed that, at points near the axis, the shock front may be regarded as a plane normal to the axis. Since the detonation process is a steady one the velocity of detonation is, therefore, by this assumption, determined entirely by the dynamical conditions within a small stream tube coaxial with the charge. For any other stream tube, off the axis, the expansion of the cross-section is, in general, greater than that of the axial tube in the same distance, and therefore leads to a smaller steady velocity. However, since the whole wave proceeds without change of form it follows that those parts of the shock front near to the cylindrical surface of the charge must be inclined to the axis. This paper, however, will not be concerned with the form of the detonation front, but only with the way in which the radius of the charge, or confinement by metal tubes, affects the velocity of detonation.

EFFECT OF THE RADIAL EXPANSION ON THE VELOCITY OF DETONATION

Consider first, from a reference system moving with the detonation wave, a stream tube, whose axis coincides with that of the charge, and which begins at the shock-wave front. Let the radius of this stream tube at any point further along the jet be denoted by r in units of its initial radius. It is convenient to work entirely with dimensionless variables such as r. Thus if p denotes the pressure in excess of one atmosphere, U the velocity of the detonation wave, Δ the loading density of the explosive, and ρ the density of the fluid at any point beyond the shock-wave front, then let

$$y = \Delta/\rho, \quad z = p/U^2\Delta. \tag{1}$$

The reaction is initiated at the shock-wave front and develops over a certain distance of the jet; this region is referred to as the reaction zone. Let the fraction of the explosive which has decomposed up to a certain point in the reaction zone be ϵ, so that ϵ varies from 0 to 1 from the shock-wave front to the point where the reaction is complete. If, at any point in the reaction zone, E denotes the internal energy per unit mass measured relative to the energy of the unexploded material at atmospheric pressure, and Q the chemical energy which has been liberated per unit mass by the reaction up to that point, then both the quantities E and Q are functions of ϵ, y and z, and each has the dimensions of the square of a velocity.

The conservation laws, applied across the discontinuity of the shock-wave front, lead, as is well known, to the equation

$$E = \tfrac{1}{2}p\left(\frac{1}{\Delta} - \frac{1}{\rho_0}\right). \tag{2}$$

The work done by an element of the fluid by expansion, per unit mass, as it passes through the detonation wave is given by $\int pd(1/\rho)$ where the integration is taken from the value of the density ρ_0 at the high pressure side of the shock front to the value ρ at the particular point under consideration. Thus, if a dimensionless function f is defined by the equation

$$f(\epsilon, y, z) = U^{-2}(E - Q), \tag{3}$$

the equation expressing the conservation of energy can be written in the form

$$f(\epsilon, y, z) = \tfrac{1}{2}z_0(1 - y_0) - \int_{y_0}^{y} z\,dy, \tag{4}$$

where z_0 and y_0 correspond to the pressure and the density at the high pressure side of the shock front, i.e. at the front of the reaction zone. In this equation it is assumed that no appreciable amount of energy is transmitted from one element of the fluid to another by heat conduction.

As has already been explained, in all cases the relative radial expansion of the stream tube which is being considered is small, and therefore it is a legitimate approximation to assume that at every point of the cross-section the pressure is the same. Also, of course, to the same approximation the density and the fluid velocity are constant over the cross-section. The fluid velocity, relative to the detonation wave at rest, is denoted by q, so that the equation of continuity is

$$qU^{-1} = yr^{-2}, \tag{5}$$

and Bernoulli's equation is

$$\tfrac{1}{2}q^2 + \int \frac{1}{\rho}\,dp = \text{constant}. \tag{6}$$

One cannot immediately carry out the integration in this equation, because here p is not merely a function of ρ, as for adiabatic expansion, but depends also upon ϵ. Using (1) and (5) and the fact that $r = 1$ when $z = z_0$, i.e. at the front of the reaction zone, (6) can be expressed

$$\tfrac{1}{2}y^2r^{-4} + \int_{z_0} y\,dz = \tfrac{1}{2}y_0^2. \tag{7}$$

Applying the laws of the conservation of mass and momentum across the shock-wave front gives, in terms of the non-dimensional variables,

$$z_0 = 1 - y_0. \tag{8}$$

Differentiating (7) one can write

$$dz = \tfrac{1}{2}y\,d(r^{-4}) - d(yr^{-4})$$

so that integrating and using (8) gives

$$z = 1 - yr^{-4} - 2\int_1^r yr^{-5}\,dr. \tag{9}$$

If the values of the variables r, y, z, when $\epsilon = 1$, are denoted by r_1, y_1, z_1, and if a mean value of y throughout the reaction zone is defined by the equations

$$\int_1^{r_1} yr^{-5}\,dr = \bar{y}\int_1^{r_1} r^{-5}\,dr = \tfrac{1}{4}\bar{y}(1 - r_1^{-4}), \tag{10}$$

then (9) becomes, for $\epsilon = 1$,

$$z_1 = 1 - y_1r_1^{-4} - \tfrac{1}{2}\bar{y}(1 - r_1^{-4}). \tag{11}$$

In the Chapman-Jouguet theory it is assumed that the gas velocity relative to the detonation wave, at the point where the reaction is complete, is just equal to the velocity of sound in the product gases at that point. This so-called Chapman-Jouguet condition has been the subject of much discussion and it now appears to be well justified both theoretically and also experimentally in so far as it leads to correctly calculated detonation velocities. This condition will also be applied here in the case where there is a small radial expansion within the reaction zone. This should, in general, be substantially correct though in the rather special case of an explosive containing a very slowly reacting component it may need modification. Accordingly

$$y_1^2\left(\frac{\partial z}{\partial y}\right)_1 + \frac{q_1^2}{U^2} = 0, \tag{12}$$

where $\partial z/\partial y$ denotes the adiabatic variation of pressure with volume. When $\epsilon = 1$, equation (4) represents the adiabatic variation of y with z so that

$$\left(\frac{\partial z}{\partial y}\right)_{\text{ad.}} = -\left\{z_1 + \left(\frac{\partial f}{\partial y}\right)_1\right\}\Big/\left(\frac{\partial f}{\partial z}\right)_1 \tag{13}$$

and the Chapman-Jouguet condition becomes, using (5),

$$\left(\frac{\partial f}{\partial z}\right)_1 = r_1^4\left\{z_1 + \left(\frac{\partial f}{\partial y}\right)_1\right\}. \tag{14}$$

When f is a known function of y and z, equations (11) and (14) determine y_1 and z_1 in terms of r_1 and $\bar{y}$. If these values are substituted into equation (4) one obtains a relation which determines U as a function of r_1. To do this first simplify (4) by using (7) and so obtain

$$f(1, y_1, z_1) = \tfrac{1}{2} - y_1 z_1 - \tfrac{1}{2} y_1^2 r_1^{-4}. \tag{15}$$

From this point no further progress can be made without some knowledge of the form of f for the product gases of the detonation. It is not an unsatisfactory approximation to assume that the internal energy is proportional to the temperature, and it will be shown that it is not necessary to know the constant of proportionality, for this can be eliminated from the final result. The dependence of f upon the pressure and the density is, however, needed and this requires an equation of state applicable to the conditions existing in the explosion products. For the purpose of calculating the detonation velocity from thermochemical data a rather complicated form is needed to give good results, but for the present purpose, which is to obtain the correction called for by radial expansion, it should be sufficient to use an equation depending simply on a constant co-volume, especially as this co-volume can be eliminated from the final result. Therefore it may be assumed that

$$f(1, y_1, z_1) = k z_1 (y_1 - b) - Q_1 U^{-2}, \tag{16}$$

where k and b are two constants, the former being related to the specific heats of the gaseous products and the latter being the co-volume in units of $1/\Delta$.

Substituting from (16) into (14), and solving the resultant equation together with (11) for y_1 and z_1, one obtains

$$(1+2k) y_1 = (1+k)\{1 - \tfrac{1}{2}\bar{y}(1 - r_1^{-4})\} r_1^4 + bk, \tag{17}$$

$$(1+2k) z_1 = k\{1 - \tfrac{1}{2}\bar{y}(1 - r_1^{-4})\} - bk r_1^{-4}. \tag{18}$$

Introducing these values of y_1 and z_1 into (15) and using equation (16) one obtains

$$2(1+2k)\frac{Q_1}{U^2} = k^2(1-b)^2 + (1+k)^2 [r_1^4\{1 - \tfrac{1}{2}\bar{y}(1 - r_1^{-4})\}^2 - 1] + k^2 b(\bar{y} - b)(1 - r_1^{-4}). \tag{19}$$

When there is no radial expansion in the reaction zone, i.e. $r_1 = 1$, the velocity of detonation may be denoted by U_0 and from (19) one obtains

$$2(1+2k)\, Q_1/U_0^2 = k^2(1-b)^2. \tag{20}$$

Eliminating Q_1 from (19) and (20) and rearranging the terms finally results in

$$\left(\frac{U_0}{U}\right)^2 = 1 + (r_1^4 - 1)\left[\left(\frac{1+k}{k(1-b)}\right)^2 \{1 - \tfrac{1}{2}\bar{y}(1 - r_1^{-2})\}\{1 - \tfrac{1}{2}\bar{y}(1 + r_1^{-2})\} + \frac{b}{(1-b)^2}\left(\frac{\bar{y} - b}{r_1^4}\right)\right]. \tag{21}$$

It is worth remarking that this equation remains valid even when $(r_1 - 1)$ is not small. The form of the equations of motion and continuity, (6) and (5), only requires that the rate of expansion along the stream tube shall be small. However, in practical applications, $(r_1 - 1)$ is usually very small compared with unity, i.e. U is, in general, only a little less than the maximum possible velocity U_0. Hence, one may, to a sufficient approximation, write $r_1 = 1$ for all terms inside the square brackets. These terms are denoted by c and therefore

$$c = \left(\frac{1+k}{k(1-b)}\right)^2 (1-\bar{y}) + \frac{b}{(1-b)^2}(\bar{y}-b). \tag{22}$$

For $r_1 = 1$ equation (17) becomes

$$k(1-b) = (1+2k)(1-y_1) \tag{23}$$

and, using this equation, (20) may be written

$$1+2k = \frac{2Q_1}{U_0^2}(1-y_1)^{-2}. \tag{24}$$

U_0 is known with considerable accuracy for many explosives, and a very good idea of the magnitude of Q_1 can be obtained from experiments such as those of Robertson & Garner (1923). These experiments do not give Q_1 exactly because the end products after adiabatic expansion are, in general, rather different from the products in the detonation wave front. The most uncertain quantity on the right-hand side of (24) is y_1. The reciprocal of this, $1/y_1$, is the compression ratio, i.e. the ratio of the density of the explosion products in the detonation front to the loading density of the explosive. Detailed calculations according to the Chapman-Jouguet theory usually give values of this ratio about 1·33, varying somewhat from one explosive to another. It will therefore be assumed that $y_1 = \frac{3}{4}$. For U_0 take $6{\cdot}8 \times 10^5$ cm./sec., which is about the value for T.N.T., and for Q_1 1000 cal./g. which is also about right for T.N.T. These values give, according to (24), $k = 0{\cdot}95$ and according to (23) $b = 0{\cdot}237$. Thus (22) becomes

$$c = 7{\cdot}24(1-\bar{y}) + 0{\cdot}41(\bar{y}-0{\cdot}237). \tag{25}$$

At the beginning of the reaction zone ρ_0 must be rather greater than the true crystal density of the explosive, since in the shock front the pressure is very high, being of order 10^5 atm. Now for cast explosives Δ is usually not very much less than the true density, so that for such explosives y_0 cannot be a small fraction. As we know from the dynamics that $y \leqslant y_1$ throughout the reaction zone, it may be assumed that for cast explosives $\bar{y}$ is approximately equal to y_1. In this limiting case, in which $\bar{y} = \frac{3}{4}$, we have, from (25), $c = 2{\cdot}02$.

For charges made up of powdered explosives the loading density may be a half, or less, of the true density, in which case $\bar{y}$ may be somewhat less than 0·75. If the value $\bar{y} = 0{\cdot}6$ is taken, then (25) gives $c = 3{\cdot}05$. Thus it must be concluded that the numerical value of the coefficient c in the formula

$$\left(\frac{U_0}{U}\right)^2 = 1 + c(r_1^4 - 1) \tag{26}$$

is approximately equal to 2·0 for cast explosives, but may be a little greater for powdered charges of low loading density.

THE RADIAL EXPANSION OF A BARE CYLINDRICAL CHARGE

In this section, the value of r_1 is calculated for a bare cylindrical charge, of radius R, detonating in a vacuum. Instead of actually finding the expansion in the reacting gases, which would be very difficult, the corresponding radial expansion in a jet of the product gases is found, assuming an adiabatic pressure volume relation. The value of r_1 is taken to be the relative expansion of the axial stream tube in a distance X equal to the length of the reaction zone. The dependence of the degree of expansion on X and the charge radius R should be very similar indeed in the two cases and, therefore, this approximation should give a satisfactory account, with equation (26), of the dependence of U upon X and R.

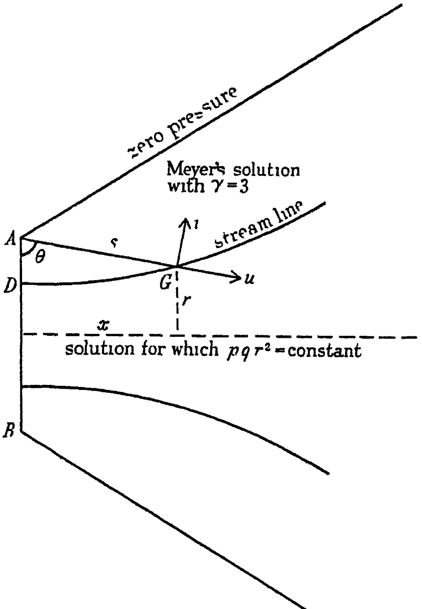


FIGURE 1

The method of calculation is best explained by reference to a figure. In figure 1 let AB denote the plane of a nozzle from which gas streams into an evacuated space from left to right. The velocity at AB is equal to the local velocity of sound. If p_0 and ρ_0 are the pressure and density at AB, is it assumed that elsewhere, in the stream, the relation between p and ρ is given by

$$p\rho^{-\gamma} = p_0\rho_0^{-\gamma}. \tag{27}$$

The appropriate value to take for γ in this equation can be obtained as follows. Consider the expansion just behind the detonation front in the purely one-dimensional case in which there is no radial expansion, then, according to the Chapman-Jouguet condition,

$$q^2 = \left(\frac{dp}{d\rho}\right)_{\text{ad.}} = \frac{\gamma p}{\rho}, \tag{28}$$

where the form (27) is used for the adiabatic relation. According to (5), for $r = 1$, it is found that $q = yU$ and therefore (28) gives $y_1 = \gamma z_1$, which together with (9), again for $r = 1$, gives

$$\gamma = y_1/(1-y_1). \tag{29}$$

As in the previous section, take $y_1 = \frac{3}{4}$, from which results $\gamma = 3$. The adiabatic relation (27) with $\gamma = 3$ will, of course, be only approximately correct for small expansions from the initial condition at AB. This, however, is all that is required in the present connexion since we are only concerned with distances along the stream less than a radius in length. The physical reason for the high value of γ is that the explosion products form initially a very imperfect gas.

The flow in the immediate neighbourhood of a point such as A or B is identical with the flow round a straight edge, which may be analyzed simply, as was first shown by Meyer. It will be assumed that the form of the stream lines, and the pressure along them, near the circular edge through A can be taken from Meyer's solution. It appears, for points not more than a radius downstream and not further from the edge than is indicated by the point D, that this approximation is not too unsatisfactory for the present purpose. Consider a tube generated by the rotation of a stream line of Meyer's solution such as DG about the axis, then within this region calculate the flow and the pressure, using equations (5) and (6). The pressure along the same line DG given by the two solutions will not, of course, be quite the same. That stream line is chosen, therefore, which minimizes the integral of the magnitude of the pressure difference along a distance R of the jet.

Let the components of the velocity at G be denoted by u and v, u referring to the component along the radius vector s, i.e. from A to G, and v the component at right angles to s counted positive in the direction of θ increasing. θ is the angle which s makes with the plane of the shock front denoted by AB. If V denotes the velocity of the issuing gas, which in this case is equal to the local velocity of sound, Meyer's solution is expressed by the following equations:

$$v = V\cos\lambda\theta, \tag{30}$$

$$u = (V/\lambda)\sin\lambda\theta, \tag{31}$$

where

$$\lambda = \{(\gamma-1)/(\gamma+1)\}^{\frac{1}{2}}. \tag{32}$$

v is given in terms of the density by the relation

$$v^2 = \frac{dp}{d\rho}. \tag{33}$$

The equation of a stream line in polar coordinates (s, θ) is as follows:

$$s = s_0(\cos\lambda\theta)^{-1/\lambda^2}. \tag{34}$$

Expressed in terms of r and θ this equation becomes

$$r = 1+\omega\{1-(\cos\lambda\theta)^{-1/\lambda^2}\cos\theta\}, \tag{35}$$

where

$$\omega = s_0/(R-s_0). \tag{36}$$

The pressure along the stream line as given by Meyer's solution is denoted by p_m. Thus, according to (27), (30) and (33),

$$\frac{p_m}{p_0} = \left(\frac{v}{V}\right)^{2\gamma/(\gamma-1)} = (\cos\lambda\theta)^{2\gamma/(\gamma-1)}. \tag{37}$$

Now consider the flow inside a tube generated by rotating the stream line (34) about the axis of symmetry. It will be assumed that inside this tube the pressure may be regarded as constant over planes normal to the axis. Thus Bernoulli's equation may be written as follows:

$$\tfrac{1}{2}q^2+\frac{V^2}{(\gamma-1)}\left\{\left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma}-1\right\} = \tfrac{1}{2}V^2, \tag{38}$$

and this, together with (27) and the equation of continuity

$$\rho q r^2 = U\Delta, \tag{39}$$

gives at once

$$r^4 = \left(\frac{p_0}{p}\right)^{2/\gamma}\left\{\left(\frac{\gamma+1}{\gamma-1}\right)-\left(\frac{2}{\gamma-1}\right)\left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma}\right\}^{-1}. \tag{40}$$

Using (35) to determine r, we find from (40) the pressure along the inside of the stream tube as given by the solution based on (38) and (39). This pressure is denoted by p_i. When $\gamma = 3$ the formulae simplify, and if non-dimensional variables are introduced as follows:

$$\xi_m = (p_m/p_0)^{\frac{2}{3}}, \quad \xi_i = (p_i/p_0)^{\frac{2}{3}}, \tag{41}$$

one gets from (35) and (40)

$$\xi_i(2-\xi_i) = \{1+\omega(1-\cos^{-2}(\theta/\sqrt{2})\cos\theta)\}^{-4} \tag{42}$$

and from (37)

$$\xi_m = \cos^2(\theta/\sqrt{2}). \tag{43}$$

The factor $(1-\cos^2(\theta/\sqrt{2})\cos\theta)$, which is denoted by $g(\theta)$, is small for all values of θ less than about 60°, in fact $g(60) = 0{\cdot}0824$. Thus, since ω is less than unity, expansion of the fourth power in (42) results in

$$\xi_i = 1-2\omega^{\frac{1}{2}}\sqrt{g(\theta)}. \tag{44}$$

Now find the stream line which gives the best approximation, i.e. the best value of ω to give the closest agreement between p_i and p_m over a distance along the axis of the order of the length of one charge radius. This is done by minimizing the integral

$$\int_0^1 (\xi_i-\xi_m)^2\,d\theta \tag{45}$$

with respect to ω.

The range of θ from 0 to 1·0 radian corresponds to a distance along the axis of about one charge radius. Minimizing (45) is simpler than minimizing the integral of $(p_i - p_n)^2$ and is as effective for the present purpose. In this way

$$2\omega^{\frac{1}{2}} \int_0^1 g(\theta)\, d\theta = \int_0^1 \sqrt{[g(\theta)]}\, \sin^2(\theta/\sqrt{2})\, d\theta \tag{46}$$

is obtained, and a numerical integration gives $\omega^{\frac{1}{2}} = 0{\cdot}922$, and therefore $\omega = 0{\cdot}85$. Denoting distance along the axis by x gives, as can be seen immediately from the figure and equation (34),

$$\frac{x}{R} = \left(\frac{\omega}{1+\omega}\right) \frac{\sin\theta}{\cos^2(\theta/\sqrt{2})}. \tag{47}$$

Hence, substituting the numerical value of ω into (35) and (47), the following equations, which give r as a function of x/R in terms of the parameter θ, are obtained:

$$r = 1{\cdot}85\,(1 - (x/R)\cot\theta), \tag{48}$$

$$\frac{x}{R} = 0{\cdot}919\left(\frac{\sin\theta}{1+\cos\sqrt{2}\,\theta}\right). \tag{49}$$

The following table shows numerically the relation between r and x/R given by these equations.

x/R	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
$(r-1)\,10^2$	0·10	0·48	1·20	2·40	4·08	6·24	8·85	11·80	15·10

The percentage diminution in the detonation velocity due to radial expansion within the reaction zone may be taken as $10^2(U_0 - U)/U$, and from (26), with $c = 2{\cdot}0$ and remembering that $(r_1 - 1) \ll 1$,

$$10^2(U_0 - U)/U = 4(r_1 - 1)\,10^2$$

approximately. Thus the percentage diminution of the velocity of detonation below its maximum value is given, for small values of $(r_1 - 1)$, by multiplying the figures of the lower line of the above table by 4. For instance, if the length of the reaction zone is one-half of the radius of the charge, the velocity will be 9·6 % below the maximum value as given by the one-dimensional Chapman-Jouguet theory, or as observed in charges of very large diameter.

It is useful, for some purposes, to define a reaction time τ by the relation $X = U\tau$, though it must be remembered, if one wants to connect τ with the actual progress of the reaction, that the velocity of an element in the reaction zone, relative to the shock-wave front, is not U but q, so that the true time of reaction is $\int_0^X q^{-1}\,dx$, which is roughly $\frac{4}{3}\tau$.

THE RADIAL EXPANSION OF A CASED CYLINDRICAL CHARGE

In this section the effect on the velocity of detonation of a metal tube surrounding the charge is considered. The present discussion is limited to a rather special case; a fuller consideration of confined charges will be given in another paper.

For the present, it will be supposed that the metal case can be represented by a thin flexible sheath of the same mass per unit length. It is not a serious approximation to neglect the strength of the metal, except in very special cases, since the yield stress, even of steel, is small compared with the stresses arising near the detonation front. However, the neglect of the finite thickness and, therefore, of the compression of the metal is a serious approximation, and it is with this that the later paper will deal.

It can be safely assumed that the expansion of the metal tube is entirely radial, so that the mass per unit length, denoted by m, is constant throughout. Moreover, for small expansions, it may be assumed that $d^2r/dt^2 = U^2 d^2r/dx^2$ so that in this case the equation of motion for the metal case may be written

$$\frac{d^2r}{dx^2} = \frac{2\pi\Delta}{m} zr. \tag{50}$$

This equation could not be solved exactly without a detailed knowledge of the way in which the rate of the reaction depends upon the temperature and pressure, for only then could we determine the dependence of z upon x and r from the equations of the first section. For small expansions, however, z may be assumed to have the same form as in the one-dimensional case. It follows that, by (9), $z = 1-y$, where y varies between Δ/ρ_0 at the beginning of the reaction zone and Δ/ρ_1 at the end. It has been seen previously that for cast explosives both these ratios are approximately $\frac{3}{4}$. Hence (50) will give a good idea of the variation of r with x, if z is taken to be constant and equal to $\frac{1}{4}$. Equation (50) may now be written

$$\frac{d^2r}{dx^2} = \left(\frac{\Delta}{4R\sigma}\right) r, \tag{51}$$

where σ is the mass per unit area of the confining tube before expansion, i.e. $2\pi R\sigma = m$. The solution of (51) for which $dr/dx = 0$ and $r = 1$ at $x = 0$ is

$$r = \cosh\left(\frac{x}{2R}\sqrt{\frac{\Delta R}{\sigma}}\right). \tag{52}$$

Thus, if X denotes the length of the reaction zone,

$$\left(\frac{U_0}{U}\right)^2 = 1 + 2c\cosh\left(\frac{X}{R}\sqrt{\frac{\Delta R}{\sigma}}\right)\sinh^2\left(\frac{X}{2R}\sqrt{\frac{\Delta R}{\sigma}}\right) \tag{53}$$

results from (26). When the percentage diminution of the velocity of detonation, due to radial expansion, is small, (53) may be written, assuming $c = 2$,

$$\frac{U_0 - U}{U} = \frac{X^2\Delta}{2\sigma R}. \tag{54}$$

As a numerical example, consider an iron tube 1 cm. internal radius and 4 mm. thick, containing an explosive at density 1·5 g./cm.^{3} and for which $X = 5$ mm. These figures give $X/R = 0{\cdot}5$ and $\Delta R/\sigma = 0{\cdot}5$ from which, by (54), the detonation velocity is found to be 6·25 % below U_0. It may be noted that the same charge completely bare would, according to the formulae of the preceding section, have a velocity of detonation 9·6 % less than U_0. Thus for such a charge the confining tube increases the rate of detonation by a little over 3 %.

The writer is indebted to the Director-General of Scientific Research (Defence), Ministry of Supply, for permission to publish this paper, and also to Sir Geoffrey Taylor for many helpful discussions.

REFERENCES

Chapman, D. L. 1899 *Phil. Mag.* **47**, 90.
Jouguet, E. 1906 *J. Math. pures appl.* **6** (Séries II), 5.
Robertson, R. & Garner, W. E. 1923 *Proc. Roy. Soc.* A, **103**, 539.

Antoine Laurent Lavoisier
26 August 1743—8 May 1794

By Sir Harold Hartley, F.R.S.

(*Delivered* 16 *November* 1944—*Received* 15 *January* 1946)

[Plates 10 and 11]

We have met to-day to honour the memory of a great Frenchman. Antoine Laurent Lavoisier is one of the immortals. In the whole history of science there is no transformation so swift and dramatic as when in 1789, in his great Treatise, he gave chemistry its modern form, sweeping away the cobwebs of centuries which obscured its progress. He was a child of his age. Bred in France at a time when the writings of Voltaire and Rousseau were stirring men's minds, and Diderot and the Encyclopaedists were widening their vision, Lavoisier was a great reformer. It is difficult within the compass of an hour to do justice to his many-track mind. Ambition, curiosity, humanity and love of action took him into many fields, and in each his creative genius saw an opportunity for constructive work. The six great volumes of his collected writings display an amazing intellectual and practical achievement. In chemistry, in physics, in physiology, in chemical engineering, in agriculture, in geology, in education, in statistics and in finance he was a pioneer, and in each his contribution has the modern touch both in thought and phrase. The marvel is that with his widespread interests it was he alone, in a band of brilliant contemporaries like Black, Cavendish, Priestley and Scheele, who had the vision of modern chemistry and gave it life. Alone among the intellectuals Lavoisier was a man of affairs, skilled in administration and finance, and that is in some measure the secret of his success. For it can be said of him that he applied not only the balance but the principle of the balance sheet to chemistry and physiology.

Lavoisier's life was such a tangled skein of occupations that they cannot be seen in their true perspective under separate headings. Their reactions on each other are the clue to so much that, even at the sacrifice of logical sequence, I shall try to picture the daily current of his life flowing in its many channels.

Lavoisier was a Foreign Member of the Royal Society and, although he never came to England, at many critical moments of his scientific work he was influenced directly or indirectly by Fellows of our Society. You will see that Robert Boyle, John Mayow, John Locke through his disciple Condillac, Stephen Hales, that most ingenious vicar of Teddington, Black, Priestley and Cavendish all had their influence on Lavoisier.

Early life

He was born in Paris on 26 August 1743. His father was a lawyer, and his mother and grandmother came of legal stock, so that he had the law in his blood, and no one could draft an opinion more skilfully. After a classical education at the

Mazarin College, he turned to science, and was fortunate in his teachers. Mathematics and astronomy he learnt from La Caille, chemistry from Rouelle, geology from Guettard, and botany from de Jussieu. From their teaching not only did he gain a broad general scientific background, but he was bitten with the idea of research.

He qualified as a lawyer, but he could afford to choose his own career, and his ambition turned to science. There was a touch of Francis Bacon in his make-up, a belief that by making systematic measurements the truth would emerge. He began like Dalton as a meteorologist, and his first published paper was on an aurora (1). Throughout his life he kept careful records of daily thermometric and barometric observations, made by himself and by correspondents in different parts of the world. He was always meaning to collate them, but that was never done, and his only meteorological paper was on forecasting, in which he suggested the value of a daily weather forecast (2). It was in meteorology that he served his apprenticeship in physical measurement, investigating the accuracy of the barometer and thermometer, and the best construction of the portable instruments which he sent to his correspondents.

GEOLOGY

In 1763, when he was twenty, Guettard interested him in his project of a geological map of France, the first of its kind. It appealed to Lavoisier's systematic mind, and he threw himself into the scheme with characteristic energy, writing to ministers for support, and drawing up questionnaires to go to the provinces. For seven years Lavoisier was making geological tours to collect data for the survey, and his journeys gave him a cross-section of social conditions all over France that left its mark on his mind.

He travelled on horse-back with his servant carrying his barometer, thermometer, hydrometer, and box of reagents. His observations included both mineral deposits and the thickness and characteristics of the various strata at different exposures. He measured in each case their height above sea-level with his barometer, so that he knew the levels at which they occurred in different places. His material formed the basis of sixteen plates in the first geological map of France which was published in 1784 by Guettard's successor Monnet without proper acknowledgement of Lavoisier's work. His only substantial geological paper, on recent sedimentary rocks (3), was not published until 1788. In it he distinguishes between littoral and pelagic banks which were formed at different distances from the land and consisted of distinct kinds of sediments and organisms. His sections gave the first outline of a correct classification of the Tertiary deposits of the Paris region. Geikie said of Lavoisier that 'if he had not given himself up to chemistry, he might have become one of the most illustrious among the founders of geology'.

PUBLIC LIGHTING

Throughout his life Lavoisier's interests were divided between pure science and its application to practical problems. He had a deep sense of responsibility that science should be used in the service of man, and he was always ready to turn aside from his researches to some practical task. In 1765 the Academy offered a prize for a paper on the lighting of the streets of a large city and Lavoisier decided to compete. He made an elaborate investigation, both practical and theoretical, of various methods of illumination; of the relative cost of lighting Paris with candle or oil lamps, of the advantage of various types of reflexion, of the time of burning of different kinds of oil, and of the use of dissolved resin to prevent the freezing of oil in winter. The prize was divided between three manufacturers, but Lavoisier's paper, an early and admirable piece of industrial research, was awarded a gold medal given by the King and was published by the Academy (4).

ACADÉMIE DES SCIENCES

1768 was the fateful year of Lavoisier's life. In February the great promise he had shown as an investigator was recognized by his election as a member of the Academy of Sciences. From the day he joined that august body as 'adjoint chimiste' his wide knowledge and experience, his diligence and his ability in drafting memoranda made him one of its most active members. Throughout his life a considerable portion of his time was devoted to its service, and Lavoisier was a most staunch defender of its privileges. He became a full member in 1778, director in 1785, and treasurer in 1791 until its suppression in 1793. The Government was constantly remitting questions to the Academy for reports, and Lavoisier's was the guiding hand in their preparation. He wrote over fifty memoranda on subjects ranging over The Great Frost of 1776, Prison Reform, Mesmerism, Water-Divining, the Creuzot Works, and The Development of Balloons (5).

THE FERME

A few days after he had joined the Academy he became a member of the Ferme, the financial corporation that leased from the Government for periods of six years the privilege of collecting the indirect taxes. The annual payments had to be made in advance, so that membership of the Ferme involved a considerable investment of capital, the return on which depended on the ability of the administration and the state of trade. For Lavoisier it was not merely an investment, as he threw himself into the task of administration with his usual energy. His practical ability and his conscientious management soon made him one of the leading members of the Corporation. Lavoisier's scientific friends shook their heads at his new venture, but said cynically 'Well, at any rate he will give us better dinners'.

The collection of customs duties, octroi, and salt tax and the sale of tobacco, involving an annual income approximating 200 million livres, required an organization of nearly 30,000 officials working all over France. This was supervised by a number of committees drawn from the sixty members of the Corporation, their work involving a series of visits to the provinces. Lavoisier served on some of these, and he was specially charged with inspection of the tobacco factories and of the frontier guards to prevent smuggling. These journeys occupied a large part of his time from 1769 to 1771.

MARRIAGE

In 1771 Lavoisier married the daughter of Jacques Paulze, one of his colleagues in the Ferme. His wife was only fourteen years old, but she was a girl of character and intelligence and devoted her life to helping her husband (see plate 10). She learnt languages and translated Priestley and Cavendish for him, and later she published a translation of the essays of Kirwan. She was constantly in the laboratory, and many pages of the records of his experiments are in her handwriting. She studied painting with David, and two drawings showing Lavoisier's experiments on respiration with Seguin are by her hand (see plate 11). She also drew and engraved all the plates for his *Traité Élémentaire de Chimie*. Benjamin Franklin, in writing to thank her for his portrait, said: 'It is allowed by those who have seen it to have great merit as a picture in every respect; but what particularly endears it to me is the hand that drew it.'

We get a glimpse of her in Arthur Young's Diary in 1787(6). 'To Monsieur Lavoisier by appointment. Madame Lavoisier, a lively, sensible, scientific lady, had prepared a *dejeuné Anglois* of tea and coffee, but her conversation on Mr Kirwan's *Essay on Phlogiston*, which she is translating from the English, and on other subjects, which a woman of understanding, that works with her husband in his laboratory knows how to adorn, was the best repast.'

Gouverneur Morris, the United States Minister in Paris, had to discuss with Lavoisier the price the Ferme should pay for Virginian tobacco and also a debt that was owing by his Government. His *Diary of the French Revolution*(7) gives us a picture of Lavoisier as a business man and of Madame Lavoisier as his hostess:

> 2 June 1789: Return Home, and then call on Monsr. Lavoisier....Talk of a future Contract. Thinks it should exist for a Part of the Supply. I agree with him and add that I would rather contract for the best Quality and leave the inferior Kinds to the Chance of Marketts. He asks at what Price. I tell him that what I say must not draw to any Consequence but be considered merely as Conversation. That it may be from 32 ₶ to 33 ₶. He thinks this much too high.
>
> 8 June 1789: Dine with Mr. de Lavoisier; as I am leaving he tells me that the Farm are determined to stand Suit and that he is sorry for it. Madame appears to be an agreable Woman. She is tolerably handsome, but from her Manner it would seem that she thinks her forte is the Understanding rather than the Person.
>
> 26 June 1789: Return Home and take a little Medicine and then go to Dinner at Mr. Lavoisier's. They are just returned from Versailles and Madame gives me the News....At Dinner, Mr. Lavoisier, who seems to be a sensible and well informed Man, tells me that the usual Produce of France is from 5 to 6 Times the Seed sown.
>
> 25 September 1789: Go to the Opera according to my Promise and arrive towards the Close of the Piece at the Loge of Madame Lavoisier....Go to the Arsenal and take Tea with Madame Lavoisier *en attendant le Retour de Monsieur* who is at the Hôtel de Ville. As Madame tells me that she has no Children I insist that she is *une Parasseuse*, but she declares it is only Misfortune....Monsieur comes in and tells us of the *Obstination* of the Bakers. This Corps threatens the Municipality of Paris with a Discontinuance of their Occupation unless a Confrère, justly confined, is released. Thus the new Authority is already trampled on. I take Mr. Lavoisier apart and propose to him to negotiate with the Farm for the Debt of the United States. I offer him also a Concern. He will I think accept it, tho he rather objects. He wishes to know my Terms. I speak vaguely.

25 July 1791: Call on Mr. Franklin and we go together to Made. de Lavoisier's to Dinner. As the Party is rather High-Flying they conclude Haphazard that the Riot at Birmingham has been occasioned by the Government. This is ridiculous enough.

7 November 1791: Mr. Franklin dines here, and we go together after Dinner to see Made. Lavoisier where there are a Number of *Gens d'Esprit* who are in general but so so Company.

Few men can have been more happily married than Lavoisier. His wife shared in all his activities, she helped him to fight his scientific battles, and in the dark days to come she risked everything to try and save him.

EARLY CHEMICAL RESEARCHES

Even before his election to the Academy Lavoisier's interests were turning from geology to chemistry. His first two chemical papers were on gypsum, its composition and its solubility in water (8). Lavoisier used a hydrometer to show that the solution of gypsum was denser than water, and this led him to a systematic study of the subject of hydrometry in which he devised a new form of constant-immersion hydrometer, a small portable version of which he used to examine the densities of a large number of natural waters during his geological tour, in order to determine their content of dissolved salts. His practical mind saw the importance of water supplies to cities, and he was a strong supporter of de Parcieux's scheme for improving the water supply of Paris, on which he read papers to the Academy in 1769.

In order to standardize his hydrometers Lavoisier studied the effect of repeated distillation on its density and also the variation of density with temperature, in which he observed the increase of density to a maximum above the freezing-point. He was unaware that this had been discovered in Italy in the seventeenth century.

In 1768 he began to investigate the supposed conversion of water into earth which had been a matter of so much controversy since the time of Boyle. Lavoisier's method of attack is characteristic of his experimental methods (9). He weighed a retort, introduced into it a known amount of fresh rain water, and luted on a closed receiver to act as a reflux condenser, after heating the retort to expel some of the air. The whole apparatus was then carefully weighed and the retort heated for 110 days in an oil-bath to about 80° C. A deposit of earth slowly formed in the retort, but the weight of the apparatus at the end was unchanged, showing that it had gained no fire material nor lost any water through the glass. This must have had an important influence on Lavoisier's ideas by strengthening his tacit assumption of the conservation of mass, and by showing him that in chemical changes involving exposure to heat the 'fire material' did not necessarily, as was often supposed, cause any change of weight. The solid was then collected, the water evaporated and the weight of the total dry solid residue was compared with that of the loss of weight of the empty retort during the operation. The solid weighed 20·4 grains, while the retort had lost only 17·4 grains. Lavoisier could not explain the greater weight of the solid residue (caused, as Meldrum has pointed out, by the absorption by alkali of carbon dioxide from the air), but he concluded rightly that in spite of this discrepancy the solid must have come from the glass and not from the water.

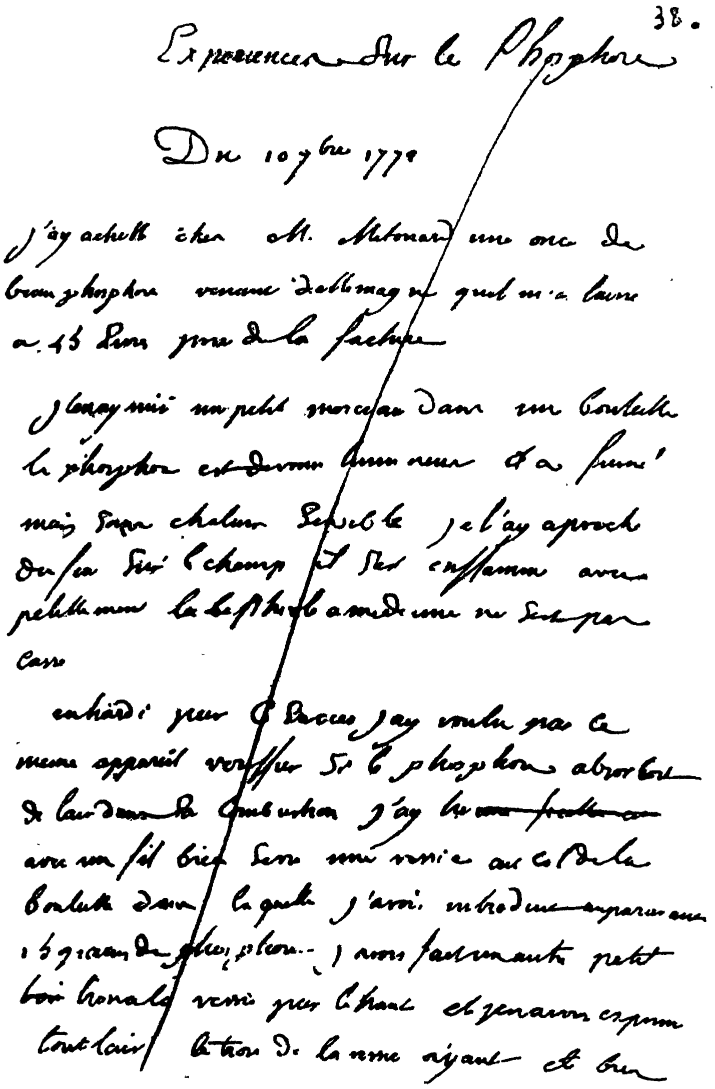
This was the most significant of Lavoisier's early papers, and it was followed early in 1772 by an investigation with Macquer and Cadet of the effect of heat on diamonds (10). The varying resistance of gem stones to heat had excited the curiosity of experimenters for centuries, and the causes of the loss of weight of the diamond on heating had been continuously a matter for inquiry. Was it due to true volatilization as Boyle thought, or to a species of combustion as Macquer's recent observations indicated, or to a decrepitation into small fragments caused by contact with cold air? Those were the rival views which Lavoisier and his friends set out to investigate by comparing the effects of heating diamonds in open and closed vessels. Heated in an open retort to a high temperature for some hours they were found to lose weight and become discoloured. On the other hand, diamonds heated in a clay pipe filled with charcoal and closed with a lute, the pipe itself being enclosed in a nest of three crucibles with a similar lute to exclude air, lost no weight and showed only a slight superficial darkening and no loss of polish after exposure to the highest temperature available. Hence they concluded that the loss of weight depends on contact with the air and is due either to combustion or to some mechanical effect of the air in causing decrepitation. These experiments were continued later by Lavoisier using the great burning glass of Tchirnhausen, but the paper was not printed until 1776, and it is not clear at what date it was presented to the Academy (11). Lavoisier's notebook shows that these experiments were still in progress in 1773.

COMBUSTION

The experiments on diamonds and probably papers by Sage and Cigna on the burning of phosphorus gave Lavoisier an interest in combustion, and his knowledge of the work of Boyle and Hales shows that he was studying the literature. At this time Stahl's theory of phlogiston still held the field. It was generally accepted by chemists that substances are inflammable because they contain this mysterious principle, which was liberated with heat and light so long as air was there to remove it. Respiration was thought to be a similar process, the air removing phlogiston from the lungs. It was a looking-glass hypothesis, a relic of alchemy, but it had an astonishing hold on chemists' minds until Lavoisier broke it.

The critical moment for Lavoisier came on 10 September 1772 (12) when, having bought an ounce of phosphorus, he found it could be burnt in glass vessels without breaking them and began to examine the absorption of air already noticed by Cigna and others during its combustion. On 20 October he sent a sealed note to the Secretary of the Academy giving the results of his experiments which showed that air is absorbed when phosphorus is burnt and that the phosphoric acid weighs more than the phosphorus (13). On 1 November he sent another note describing his discovery that sulphur, like phosphorus, gains in weight during combustion, and that this increase of weight is due to a prodigious quantity of air which is fixed during combustion (14).

Cette découverte, que j'ai constatée par des expériences que je regarde comme décisives, m'a fait penser que ce qui s'observait dans la combustion du soufre et du phosphore pouvait bien avoir lieu à l'égard de tous les corps qui acquièrent du poids par la combustion et la calcination;...L'expérience a complétement confirmé mes conjectures; j'ai fait la réduction de la *litharge* dans les vaisseaux fermés, avec l'appareil de Hales, et j'ai observé qu'il se dégageait, au moment du passage de la *chaux* en métal, une quantité considérable d'air, et que cet air formait un volume mille fois plus grand que la quantité de *litharge* employée. Cette découverte me paraissant une des plus intéressantes de celles qui aient été faites depuis Stahl, j'ai cru devoir m'en assurer la propriété, en faisant le présent dépôt entre les mains du secrétaire de l'Académie, pour demeurer secret jusqu'au moment où je publiera mes expériences.



38.

Experiences sur le Phosphore

Du 10 7bre 1772

J'ay acheté chez M. Mitouard une once de beau phosphore venant d'allemagne qu'il m'a laissé a 15 livres prix de la facture

J'en ay mis un petit morceau dans une bouteille le phosphore est devenu lumineux et a fumé mais sans chaleur sensible je l'ay aproché du feu sur le champ il s'est enflammé avec petillement la [illegible] ne s'est pas cassé

enhardi par ce succes j'ay voulu par le meme appareil verifier si le phosphore absorboit de l'air dans sa combustion j'ay lié ~~[illegible]~~ avec un fil bien serré une vessie au col de la bouteille dans laquelle j'avois introduit ~~auparavant~~ 15 grains de phosphore j'avois pareillement un petit bois [illegible] vers le haut et j'en avois exprimé tout l'air le tour de la vessie ayant été bien

FIGURE 1. The page from Lavoisier's notebook, 10 September 1772, recording his experiments in burning phosphorus.

It was a moment of great elation for Lavoisier. To his practical mind increase in weight meant gaining something, not losing it, as the phlogistics argued. He saw a vision of something much bigger than a theory of combustion, 'a revolution in chemistry and physics'. He saw, too, that its realization depended on a detailed investigation of the gases that were given off or absorbed in the chemical changes. It was no easy task, as Lavoisier had had comparatively little chemical experience, but he attacked it with courage on a wide front, and the next twelve months saw the most concentrated and sustained scientific work of his career. He was confident of his objective, and that confidence kept him going through the perplexities and disappointments of twelve crowded years.

OPUSCULES PHYSIQUES ET CHIMIQUES

The record of his next year's work was published in a separate volume, *Opuscules physiques et chimiques*, as it was too long for the *Memoirs* of the Academy (15). It is described in the preface as the first of a series of investigations covering the wide field that he already had in view. It was Lavoisier's apprenticeship to the study of gases. He began with a careful study of the literature, and half the volume consists of a history of previous work. Hales, Black and Priestley are the three chemists to whom he pays special attention. Through Hales he learned the experimental methods of John Mayow, and his own apparatus, though more elaborate, is obviously based on that of Hales. His systematic mind appears in the tables in which he assembles all the quantitative results of Hales's ingenious but almost random experiments. From Black he learned perhaps more than he said. Black's paper on the alkalis was a quantitative study after his own heart, and broadly speaking he was to do for metals and oxides just what Black had done for mild and caustic alkalis.

Having finished his survey of the literature, on 22 February 1773 he began his laboratory notebook with a remarkable forecast of the work it was to record (16):

> Qui m'a paru fait pour occasionner une révolution en physique et en chimie. J'ai cru ne devoir ne regarder tout ce qui a été fait avant moi que comme des indications: je me suis proposé de tout répéter avec de nouvelles précautions, afin de lier ce que nous connaissons sur l'air qui se fixe, ou qui se dégage des corps, avec les autres connaissances acquises et de former une théorie.

Lavoisier's own experiments, which are described in the second half of the volume, consist mainly of careful measurements of the changes in weight and volume that accompany the evolution or absorption of gas in the formation or decomposition of carbonates, the neutralization of acids, the reduction of metallic oxides, and the calcination of metals. His observations were thus concerned only with carbon dioxide and oxygen, but at that time Lavoisier had formed no definite view as to the nature of different gases. He was convinced of their fundamental importance and his aim was to bring them all into a common theory.

The experimental part of the paper begins with a repetition and testing of Black's work, using the results to establish the composition of quicklime, slaked lime and chalk. This was then extended to the carbonates and hydroxides of the alkali metals and ammonium, and to the precipitation of metallic solutions by alkalis. The amounts of gas evolved by heating metallic calces with charcoal were measured and the gas was shown to be identical with that from chalk, but owing to faulty experimental methods the gas in each case was mixed with air, which must have complicated Lavoisier's conclusions. Finally, he measured the absorption of air when metals are calcined or phosphorus burnt in bell-jars over water.

Lavoisier's most important conclusion was that in confined spaces only a limited amount of calcination takes place, the increase in weight being roughly proportional to the volume of air absorbed. Thus it followed that metals could only fix part of the air, and the same result was obtained with phosphorus. But no conclusion is hazarded as to the nature of the gas which supports combustion, and the last sentence of the paper reveals one of the difficulties that must have been puzzling Lavoisier's mind. He says:

> J'ai été curieux, relativement à des vues dont je rendrai compte dans un autre temps, d'observer si le mélange d'un tiers de fluide élastique des effervescences corrigerait l'air qui avait servi à la combustion du phosphore, et lui rendrait la propriété d'entretenir les corps enflammés. Le mélange fait, j'en ai rempli un bocal étroit et j'y ai introduit une bougie, mais elle s'y est éteinte sur-le-champ.

The paper as a whole is such a clear exposition of a logical sequence of quantitative experiments that it is disappointing that it should end on so inconclusive a note. Luckily we have in Lavoisier's notebook the clue to his perplexities. Misled perhaps by Black's name of 'fixed air' for carbon dioxide (which he abandoned in the following year), Lavoisier thought that the gas which was 'fixed' in alkalis was identical with the gas 'fixed' by metals in their combustion to form calces. This is proved by several entries in his notebook; for example, on 1 July 1773 he wrote: 'Persuadé que la combustion du phosphore absorbe l'air fixe contenu dans l'air, ou plutôt le soupçonnant, j'ay pensé qu'en rendant de l'air fixe à cet air, on pourrait peut-être le rendre air commun.' He tries, of course, without success, but fails to push the matter further (17).

A comparison of the notebook with the paper is interesting in another way. The experiments seem to have been made without any planned sequence, but in the paper the results are marshalled in perfect order, showing the systematic grasp Lavoisier already had of the whole field, his doubts only being recorded in the last sentences. It was a remarkable pioneer achievement in experimental technique, but progress had to wait until Lavoisier knew the relationship between carbon dioxide and oxygen. He had still much to learn, despite his brilliant intuition.

CALCINATION OF METALS

Lavoisier was unshaken in his belief that increase in weight meant chemical combination, and immediately the *Opuscules* was published he began experiments on the calcination of tin and lead in sealed vessels. Boyle had done similar experiments but had only reweighed the vessels after opening them, when he found an increase in weight which he explained by the passage of fire material through the glass (18). Lavoisier saw the opportunity for a crucial experiment by weighing the vessels both before and after they were opened. A summary of his results was deposited with the Academy on 24 April 1774, but was not read until the public session on 14 November (19), following Lavoisier's usual plan of keeping a tit-bit for that occasion. Without giving any numerical details Lavoisier said that in sealed vessels the calcination of lead and tin ceased after an hour's heating, the amount of calx formed being greater in larger vessels. In every case the weight of the vessel was unaltered if it was weighed before it was opened. When the seal was broken, air was heard to rush in, and the weight of the vessel was then greater, the increase being 'exactly proportional to its capacity'. Hence it was clear that the increase in weight was due not to the passage of fire-stuff through the glass but to something the metal had 'borrowed' from the air in the vessel which had converted it into a calx.

Lavoisier ended with a significant sentence about the nature of the air that remained after metal had been calcined in it:

> Cet air ainsi dépouillé de sa partie fixable (je pourrois presque dire de la partie acide qu'il contenoit); cet air, dis-je, est en quelque façon décomposé; et il m'a paru résulter de cette expérience un moyen d'analyser le fluide qui constitue notre atmosphère; et d'examiner les principes qui le constituent. Quoique je ne sois pas arrivé à cet égard à des résultats entièrement satisfaisans, je crois cependant être en état d'assurer que l'air aussi pur qu'on puisse le supposer, dépouillé de toute humidité et de toute substance étrangère à son existence et à sa composition, loin d'être un être simple, un élément, comme on le pense communément, doit être rangé au contraire tout au moins dans la classe des mixtes, et peut-être même dans celle des composés.

Lavoisier found that his experiments were not as new as he had thought, and he added a note saying that Beccaria had made the same observations on tin fifteen years earlier, and that Priestley had pointed out the limited extent of calcination in sealed vessels. But the ultimate value of an observation lies in the deduction from it, and here Lavoisier alone saw the truth.

A full account of these experiments was not given until 1777, when Lavoisier had to admit that owing to their difficulty only two with tin were completely successful and none with lead (20). However, the numerical results of the two successful experiments showed the care and skill with which they had been carried out, and justified the conclusion that the weight of the sealed vessels was unaltered, within the errors of experiment, after calcination had occurred. By then Lavoisier had further evidence about the nature of air, which he describes as a mixture of a salubrious portion which combines with metals during calcination, leaving 'une espèce de mofette' which will not support life or combustion, and is itself 'fort composé'. From his results he suspected that the former is the denser of the two, air being intermediate. His experiments were, in fact, sufficiently accurate to justify this conclusion.

These experiments on the calcination of tin must have strengthened Lavoisier's confidence in the correctness of his theory, and from March 1774 onwards his notebook shows that he was widening his front of attack, for combustion was only one aspect of the revolution he had in mind (21). This year saw the beginnings of work on a number of problems whose solution eluded him for the moment, as the background of his knowledge was as yet insufficient. In each there was a quantitative approach to determine both the composition of substances and the nature of the chemical changes that they undergo.

He begins by burning inflammable air (hydrogen) which ought on his theory to increase in weight, and he was puzzled to find a loss owing to the escape of water vapour. This was a problem that continued to perplex him for nine years.

Next came a study of the formation of nitrous ether, also abortive, then an attempt to determine the composition of nitre by exploding a mixture of it with charcoal, and analyzing the gaseous products. Knowing the percentage of caustic alkali in the nitre he tried to calculate its composition, but he did not know how much the charcoal contributed to the weight of fixed air.

In October he is repeating a number of Priestley's experiments on various gases, including his production of an acid by passing electric sparks in air. But it was characteristic of Lavoisier's lack of qualitative intuition that in spite of his interest in acids he does not follow up this clue, and the explanation had to wait for Cavendish ten years later.

Next come some revealing thoughts on 'vegetable analysis' and plans for future work, the birth of organic analysis.

> Nous ignorons: 1° Quelle est la qualité de cette immense quantité d'air qui se dégage pendant la distillation: il y a probablement de l'air fixe et de l'air inflammable; 2° ce que c'est que l'huille: il paraît que par la combustion on peut la réduire en air et en eau; mais nous ne savons rien au de-là. Déterminer les proportions des deux en brûlant une lampe en un vase clos. 3° Ce que c'est que le charbon. Nous savons bien qu'en brûlant il convertit l'air évidemment en air fixe: mais nous ne savons pas s'il donne lui-même de l'air fixe.

Lavoisier is still not quite free from the incubus of phlogiston, and he suggests burning charcoal in a sealed vessel to see if any phlogiston escapes through the glass during combustion with a corresponding loss of weight.

PRIESTLEY AND OXYGEN

Lavoisier's notebooks show that in 1774 his mind was roving over a wide range of problems of great significance but too complex to be solved with his present knowledge. He was still puzzled about the relation of fixed air to the air absorbed in calcination, and had tried to find a calx which could be reduced and give up this air in the absence of charcoal. It seems clear that Priestley gave him the clue he needed when he was in Paris in 1774. On 1 October he told Lavoisier that by heating the calx of mercury he had got a gas in which a candle burnt particularly brightly. It was not until the following 1 March that Priestley realized the remarkable properties of this gas and thus discovered oxygen. Lavoisier got some calx of mercury from Cadet, and after some preliminary experiments in November three months elapsed until he completed them in three days' work with Trudaine at Montigny. He described the results at the public session of the Academy in April 1775 in a paper on 'The nature of the principle that combines with metals during their calcination and augments their weight' (22).

His main interest was to compare the fixed air he got by heating the calx of mercury with charcoal with the gas he got by heating it alone, which to his surprise was quite different and 'more respirable, more combustible and consequently purer than ordinary air'. His difficulty about the relation of these two gases was therefore solved. But qualitative experiments were never Lavoisier's strong suit, and in this paper for once he is inconsistent. Perhaps it was written hurriedly for the public session. At one point he describes the 'principle' absorbed in calcination as 'neither one of the constituents of the air, nor a particular acid existing in it, but the air itself without alteration or decomposition'. And he has the curious idea that the air can be purified whenever it is absorbed in a calx and then set free again. In another part of the paper he says that the 'principle' is the purest part of the air. Lavoisier is rarely guilty of such inconsistency, but in this paper he clearly failed to realize the full significance of the facts he had observed.

The paper was published in Rozier's *Journal* and soon drew a criticism from Priestley (23), whose observations were much more accurate:

> Having mentioned the paper of Mr Lavoisier's,...I would observe, that it appears by it, that, after I left Paris, where I procured the *mercurius calcinatus* abovementioned, and had spoken of the experiments that I had made, and that I intended to make with it, he began his experiments upon the same substance, and presently found what I have called *dephlogisticated air*, but without investigating the nature of it, and indeed, without being fully apprised of the degree of its purity....He therefore inferred, as I have said that I myself had once done, that this substance had, during the process of calcination, imbibed atmospherical air, not in part, but in whole....As a concurrence of unforeseen and undesigned circumstances has favoured me in this inquiry, a like happy concurrence may favour Mr Lavoisier in another; and as, in this case, truth has been the means of leading him into error, error may, in its turn, lead him into truth.

It is curious that Lavoisier failed to see at once that the experiments with the calx of mercury gave the support he needed for his theory, and that he was so slow to complete them. From May until September 1775 his laboratory notebook is a blank. The reason for all this was a new preoccupation in his mind.

Régie des poudres

Turgot, a friend of Lavoisier, had just been made Controller-General, the one man who might perhaps have saved France from a Revolution, if vested interests had not been too strong. Lavoisier was in touch with every side of the administration, and his critical eye had noted the inefficiency of the national production of gunpowder which was made under contract by a financial company. Lavoisier saw that their methods of collecting nitre and of making gunpowder were costly and wasteful, and that production was insufficient. He remembered, too, that the prohibitive cost to France of imported powder during the Seven Years' War was one cause that had led to the peace of 1763. He succeeded in persuading Turgot that it would not only pay the government to cancel the existing contracts with compensation and take over the manufacture itself, but that it was essential for the national safety. This was done by a decree in March 1775, and Lavoisier and three of the former officials were put in charge of the powder factories.

Lavoisier at once concentrated all his energies on the new task. He moved to the Arsenal, where he lived and had his laboratory until 1792. He then began a thorough investigation into the methods of making and collecting nitre, nitrification being the only source of saltpetre apart from some deposits which he spent three months in surveying. His laboratory notebooks for 1775 and 1776 are full of records of various investigations into the manufacture of nitre and gunpowder.

In 1777 he drew up comprehensive instructions for the construction and operation of nitrification plants embodying many improvements, which represented a fine piece of chemical engineering (24). With his usual eye to finance he calculated that the plants should yield a net revenue of 15 % on the capital. Great care was taken to select suitable staff, who had to pass an examination in chemistry, mathematics and the construction of powder mills.

The results were even better than Lavoisier had predicted. In three years the improvement of the powder had increased the range of French weapons by 60 %. Production increased steadily, and by 1788 there was a reserve of five million pounds in the magazines, and the economies had amounted to 20 million livres.

However, the effect of Lavoisier's work did not end with the Napoleonic wars. He had interested young Irénée Du Pont, the son of the physiocrat, in chemistry, and later he became Lavoisier's assistant at the Arsenal. Irénée and his father were constitutionalists, and in 1792 they were fighting beside the Swiss Guard at the Tuileries. Saved by a miracle, they emigrated later to America. There on the banks of the Brandywine River at Wilmington, Irénée Du Pont built another powder factory destined twice to become a great arsenal of the Allied Nations.

THE DISCOVERIES OF 1776 AND 1777

In September 1775 experiments in the laboratory were resumed, though most of the work in the autumn was concerned with the manufacture of nitre. But gradually Lavoisier began to take up his scientific work again. He dissolves mercury in nitric acid, evaporating the solution to dryness and measuring the gases evolved when the mercuric nitrate is heated. This he uses as a method of analyzing nitric acid, as the experiment ends with the same weight of mercury as at the start. From the proportions of nitrous gas and pure air he calculates the percentage composition of nitric acid, assuming the other component to be water. The results were read to the Academy in 1776 in a paper 'On the existence of air in nitric acid' (25), which he takes as an illustration of his general thesis that 'non-seulement l'air, mais encore la portion la plus pure de l'air, entre dans la composition de tous les acides sans exception; que c'est cette substance qui constitue leur acidité'. He shows how he gets a mixture of nitrous air and pure air from nitric acid, and how by recombining them he gets nitric acid back again. He always tried to verify his conclusions both by analysis and synthesis.

In this paper he twice pays a tribute to Priestley as the original discoverer of many of the facts he describes: 'comme les mêmes faits nous ont conduits à des conséquences diamétralement opposées, j'espère que, si on me reproche d'avoir emprunté des preuves des ouvrages de ce célèbre physicien, on ne me contestera pas au moins la propriété des conséquences'.

This was the only paper he published in 1776, but it was a very busy year in the laboratory. Many of Priestley's experiments were repeated. Work was in progress on the analysis of air by the calcination of mercury, on the composition of various acids, on respiration, on the burning of a candle, on the latent heat of vaporization, and on the heat produced in the neutralization of acids.

Lavoisier is now quite clear about the nature of the air and of combustion, and for the first time he openly challenges the theory of phlogiston. During 1777 the evidence he had accumulated in support of his theory was presented to the Academy in nine papers, one of which was not read until 1779, and none was printed until 1780. Air, he says, is a mixture, four-fifths being an inert gas 'mofette' which takes no part in respiration or combustion, and one-fifth a gas 'eminently respirable', a constituent of all acids, which he therefore named oxygen. The compositions of carbonic acid, nitric acid, phosphoric acid, sulphuric and sulphurous acids had been determined. Respiration and the burning of a candle both convert oxygen into carbonic acid or 'acide crayeux aeriforme', the name Lavoisier has substituted for fixed air. Slow combustion in the lungs is suggested as the source of animal heat. Finally, the existence of the three states of matter is shown to depend on their heat content, and the heat of combustion is explained by the heat given out by gaseous oxygen in its change of state and combination.

Supported by this evidence Lavoisier attacks the theory of phlogiston, and shows that his own explanation is much simpler, involving no unwarranted assumptions such as the presence of an immense quantity of fire material in solids like the diamond, or that substances gain in weight while losing part of their contents.

It was a year of great achievement; the problem of combustion was solved, but Lavoisier's theory gained no adherents. It was incomplete, and there were many other simple reactions he could not explain. He was still puzzled by inflammable air, and he had not tried to discover the nature of the inert constituent of the atmosphere.

LAVOISIER'S EXPERIMENTAL FARM

The clue to both problems was to come eventually from Cavendish, and Lavoisier's notebooks of 1778–82 contain no records of new and important discoveries. His hands must have been full enough with the Arsenal, the collection of taxes and the business of the Academy. To these he now added a fresh activity—an experimental farm. On his many journeys through France Lavoisier, like Arthur Young, had seen the poor state of French agriculture and the hard lot of the peasants. The staple crop was wheat grown three years out of four, and the head of livestock was too small to yield the manure required for good cultivation. Lavoisier saw the advantages of the British system of mixed farming, with its larger capital investment, more livestock, and wheat grown only one year in four.

So in 1778 he started an experimental farm at Fréchines near Blois, in a district where the standard of farming was low, in order to investigate and demonstrate the possibilities of improvement. Lavoisier always took the big view, and his aim now was to increase the wealth of France and to ease the life of the peasants.

The experiments were carried out as far as possible with the same strict control as if they had been made in the laboratory. He kept records at Paris of every field, its size, soil, crop and yield, with an annual debit and credit account.

His main objective was to increase the amount of fodder crops and raise the number and quality of his livestock. He had some disappointments at the start—lucerne failed and clover did badly in dry years, but sainfoin was a success. He introduced ley farming on the fallows, catch crops and root crops, and gradually increased his livestock. The folding of sheep was another successful innovation. No point was too small for Lavoisier, and one of his papers describes in great detail a simple way of making hurdles.

After nine years' development the farm was in much better shape, but Lavoisier was disappointed at the slowness of progress, and he saw no prospect of getting a 5 % return on his investment (26). This and the shortness of the leases explained why French farming suffered so badly from a lack of capital. However, that was not the end, as by 1793 his wheat crop had doubled, his stock had multiplied five times, and his neighbours were copying his improved methods of farming.

When Calonne appointed a Consultative Committee on Agriculture in 1785, Lavoisier was its secretary and drafted its reports. Among its schemes were projects for experimental farms, a museum of agricultural implements, and a school of textiles made from home-grown materials, which was actually started in Paris. Among Lavoisier's friends were the leading physiocrats, including Du Pont de Nemours and Malesherbes. While he sympathized with their views on freedom of trade, and regarded agriculture as the primary source of national wealth, he recognized that industry, too, had its place in the national economy. One of his memoranda dealt with the bad effects of restrictive regulations on agriculture. Another, dealing with the many relics of feudalism which handicapped the French farmer, reads like pages from Arthur Young's Diary. A third, 'Instructions to the Provincial Assemblies for improving agriculture', recommends many of the new practices Lavoisier had tried himself at Fréchines. In the end, however, little was done owing to the lack of interest of the ministers.

LAVOISIER'S DAY

Lavoisier's immense output of work in so many fields during this period was only made possible by his quickness of mind and memory, his power of concentration on the matter in hand and his methodical habits. While he was living at the Arsenal six hours a day were given to science, in the early morning from 6 o'clock to 9 and from 7 o'clock to 10 after dinner. The rest of the day was spent in dealing with the business of the Ferme or the powder factories, in the meetings of the Academy or of the many Commissions of which Lavoisier was a member. Sunday was the happiest day of the week, as he spent the whole of it in his laboratory, which soon became a rendezvous for intellectual society in Paris. Scientists, ministers, economists and distinguished visitors from abroad went there to see Lavoisier's latest experiment and to join in the discussion of its significance. This was the usual preliminary to one of his papers at the Academy. The younger men interested in science were equally welcome with their seniors, and they often helped Lavoisier with his experiments. The memories of those brilliant gatherings in his laboratory, which played no unimportant part in scientific history, remained fresh in the minds of those who shared in them.

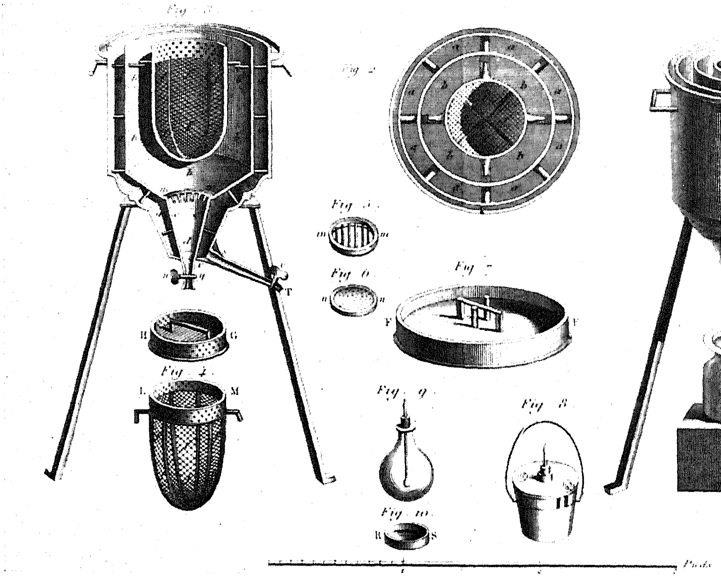
THERMOCHEMISTRY

One of Lavoisier's great services to chemistry was to disentangle the chemical and physical aspects of chemical change. The phlogiston theory had been responsible for much loose thinking, since it assumed that this mythical principle could modify both the chemical and the physical properties of matter. Lavoisier concentrated first on the alterations in weight accompanying chemical change, particularly combustion. His belief in the conservation of matter was strengthened by his experiments in sealed vessels which showed no evidence of changes in weight due to the gain or loss of fire material. This was confirmed by unpublished experiments which proved that the weight of sealed tubes of water remained unchanged when they were frozen. The distinction between chemical and physical changes simplified his task greatly, and by 1777 his evidence on the gravimetric side was sufficient to enable him to give a satisfactory explanation of the nature of air and its role in the chemical changes involved in combustion and respiration.

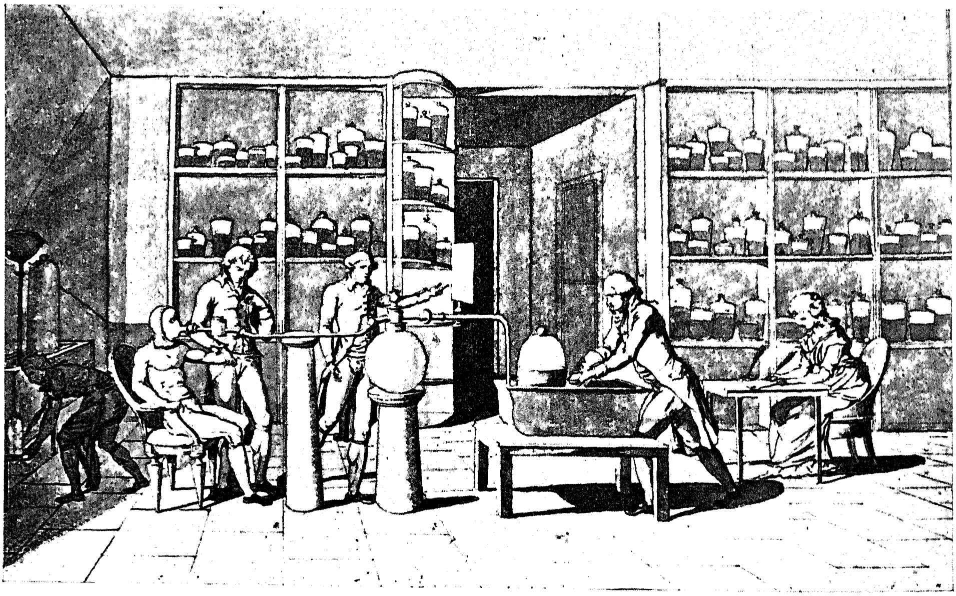
Lavoisier was now striving to give a similar picture of the physical changes accompanying chemical action. He recognized the dependence of the three states of matter on their heat content, and he explained the heat of combustion by the change in the heat content of the oxygen concerned. He had no idea of the conception of energy, and he did not connect the evolution of heat with the affinities of the elements for each other. The difference in the heat of neutralization of caustic and mild alkalis (hydroxides and carbonates) he explained by the absorption of heat in the liberation of carbon dioxide as a gas.

Lavoisier had always been interested in thermometry and heat changes since the start of his scientific work. In 1772, when Desmarest told the Academy of Black's

Portrait of Lavoisier and his wife by David, 1788.



Lavoisier and Laplace's ice calorimeter.



Madame Lavoisier's drawing of an experiment on respiration; Seguin is wearing a mask, Lavoisier directing the experiment, and Madame Lavoisier taking notes.

work on latent heat, Lavoisier described an experiment made the preceding year in which he had discovered independently the same phenomenon, finding to his surprise that on mixing water and crushed ice the temperature of the water remained at zero until all the ice was melted (27). In 1773, when studying the crystallization of sodium sulphate by cooling solutions, he observed the temperature arrest which occurred when crystallization began and lasted until it had finished. This he explained correctly by the heat of crystallization of the salt (28).

In his physical work he had the advantage of partnership with Laplace, who worked in his laboratory at intervals from 1777 to 1785. In 1777 they were measuring the vapour pressure of liquids, and determining latent and specific heats by the method of mixtures. Black's unpublished work was known to them, and de Luc and Crawford were working in the same field. In 1782, finding the method of mixtures unsatisfactory for measuring heats of reactions, they invented a new method, the ice calorimeter, in which the amount of heat given out is measured by the amount of ice melted (plate 11). The calorimeter consisted of two concentric vessels each containing crushed ice, the purpose of the outer vessel being to insulate the inner one against any gain or loss of heat from outside. There was a space in the inner vessel to hold the substance or reaction under examination, the amount of heat enclosed being measured by the amount of water that ran down from the ice as it melted. It was a beautifully simple and accurate apparatus which they used to measure specific heats, including the specific heats of gases by a flow method, heats of reaction, heats of combustion of phosphorus, carbon, ether and oil, and the heats of detonation of nitre mixed with charcoal or sulphur.

Thermochemistry dates from Lavoisier and Laplace's great paper 'Sur la Chaleur' of 1783 (29), and one of their generalizations is a partial anticipation of Le Chatelier's Theorem which came a century later: 'Dans les changements causés par la chaleur à l'état d'un système de corps, il y a toujours absorption de chaleur; en sorte que l'état qui succède immédiatement à un autre, par une addition suffisante de chaleur, absorbe cette chaleur sans que le degré de température du système augmente.'

More measurements were made in the following winter, including the heats of combustion of hydrogen and of wax, which Lavoisier showed was nearly equal to the sum of the heats of combustion of the hydrogen and carbon it contained.

The ice calorimeter thus gave Lavoisier precise data for the heat changes involved in the chemical reactions in which he had already determined the weight changes involved. He now had a complete picture of these two different aspects which enabled him to show how unnecessary it was to assume the existence of phlogiston, as everything could be explained much more simply without its aid.

In 1777 Lavoisier had explained the source of animal heat as being the heat evolved when carbonic acid (*air crayeux*) is produced in the lungs from the oxygen of the air during respiration. This was an entirely new point of view, as Haller had accepted Stahl's explanation that the heat of the body was due to the friction of the blood in the arteries, though he was baffled by the constancy of body temperature. The ice calorimeter gave the opportunity to verify this speculation by measuring the heat given out by a guinea-pig and comparing it with the heat evolved in the formation of the volume of carbonic acid which an animal of similar weight expired during the same period.

First the amount of carbonic acid expired by guinea-pigs in a period of ten hours was measured by absorbing it in potash bulbs. From the mean of their experiments Lavoisier and Laplace calculated that the combustion of the corresponding amount of carbon would have melted 10·38 oz. of ice. A guinea-pig was then kept in a calorimeter for the same period and the heat given out by it melted 13 oz. of ice. Part of the excess, they said, might have been due to the cooling of its extremities—it might well have had cold feet! From the results they concluded: 'la conservation de la chaleur (animale) est due, au moins en grande partie, à la chaleur que produit la combinaison de l'air pur respiré par les animaux avec la base de l'air fixe que le sang lui fournit'. The maintenance of a constant body temperature under different climatic conditions they ascribed to differences in the rate of evaporation of moisture rather than to differences in the amount of carbonic acid formed during respiration.

This was one of Lavoisier's most brilliant papers, remarkable for the beauty of the experimental technique, for the directness of attack, and for its verification of his theory of animal heat, which gave a new significance to metabolism. No doubt the paper owed much to Laplace, who was the first convert to Lavoisier's theories. The full significance of their eight years of partnership has not had the attention it deserves.

The nature and composition of water

His notebooks show repeatedly how puzzled Lavoisier still was as to the nature of the inflammable air he got in various ways, by the solution of metals in acids, or the distillation of vegetable substances. In 1774 he had tried and failed to ascertain the increase in weight when inflammable air was burnt. He expected inflammable air to give an acid on combustion but could find none. Since the inflammable air came from sulphuric acid and metals Lavoisier thought it should give sulphuric acid. Bucquet thought it should give fixed air. Experiments in 1774 and 1777 showed that neither acid was produced. He returned to the problem again in 1781 and 1782, when he and Gingembre burnt a jet of oxygen in inflammable air and found neither carbon dioxide nor any acidity, which still surprised Lavoisier 'que l'analogie m'avait porté invinciblement à conclure que la combustion de l'air inflammable devait également produire un acide'. Hence in 1782 he had no idea that water was the product.

Meanwhile other chemists were busy with the same problem, and Cavendish had found that when a mixture of inflammable air and common air is exploded in a closed vessel water is formed without any change in weight. When Blagden, the Secretary of the Royal Society, was in Paris, he told Lavoisier of Cavendish's experiments. On 23 June 1783, Lavoisier took two gas holders containing inflammable air and oxygen and joined them up to a jet at which the mixture could be burnt for a considerable time in a bell-jar over mercury. The first experiment was made by Lavoisier and Laplace in the presence of Blagden and a number of other scientists. A considerable quantity of pure water collected on the surface of the mercury, and although it was not possible to establish directly that the weight of the water found was equal to that of the two gases Lavoisier had no doubt of it. He wrote 'comme il n'est pas moins vrai en physique qu'en géométrie que le tout est égal à ses parties, de ce que nous n'avions obtenu que de l'eau pure dans cette expérience, sans aucun autre résidu, nous nous sommes crus droit d'en conclure que le poids de cette eau étoit égal à celui des deux airs qui avoient servi à la former' (30).

The next day the experiment was described to the Academy. 'Nous ne balançâmes pas à en conclure que l'eau n'est point une substance simple, et qu'elle est composée poids pour poids d'air inflammable et d'air vital.' Lavoisier then heard that Monge had made similar experiments. Having determined the combining volumes of the two gases he had found that the weight of the water was almost equal to that of the two gases, the specific gravities of which he had determined.

Lavoisier must have seen in a flash the wide significance of this new view of the nature of water, which gave him the clue to so many of his unsolved problems. It explained the source of the water he got in organic combustions, and he saw the possibility of using them to determine the composition of organic bodies.

Additional evidence was obtained by studying the decomposition of water by long standing over cold iron, and by passing steam over hot iron or charcoal, as well as from experiments on its formation by passing inflammable air over heated metallic calces, thus extending some observations made by Priestley.

Lavoisier was naturally anxious to know the percentage composition of water, as it would enter into so many of his calculations. In his first paper to the Academy he gives the following figures based on experiments made with Meusnier to determine the combining volumes of the two gases and their densities:

	Livres
Air vital, ou plutôt principe oxygine	0·86866273
Air inflammable, ou plutôt principe inflammable de l'eau	0·13133727
Total	1·00000000

Lavoisier at this stage has a preference, like Boyle, for the word 'principle' rather than 'element'. The name of hydrogen was not adopted until the whole nomenclature was revised four years later.

There are numerous references to a new apparatus that was constructed to prove that the weight of water was equal to that of the gases from which it was formed, and to determine its composition. Arthur Young saw it in Lavoisier's laboratory in 1787. 'In the apparatus for aerial experiments', he writes in his Diary (31) 'nothing makes so great a figure as the machine for burning inflammable and vital air, to make or deposit water; it is a splendid machine. Three vessels are held in suspension with indexes for marking the immediate variations of their weights; two that are as large as half hogsheads contain the one inflammable, the other the vital air, and a tube of communication passes to the third, where the two airs unite and burn;....If accurate (of which I confess I have little conception) it is a noble machine.'

It has been questioned whether Lavoisier ever used this apparatus successfully, as the results were never published by him. The experimental details, however, are given in the *Traité*. Hydrogen was burnt at a fine jet in a glass globe containing oxygen, both gases having been dried by passing over potash (Lavoisier says that calcium nitrate or chloride would have been better). The increase in weight of the globe together with measurements of the volumes of the gases consumed, and a knowledge of their densities, gave Lavoisier the data he needed. Details of one experiment were published in an unsigned article in the *Journal Polytype* (32), an ephemeral scientific journal in which Lavoisier was interested. The weight of water found was within 1 % of that of the two gases, and its composition is given as 85 % oxygen and 15 % hydrogen, which are the figures Lavoisier used subsequently in all his calculations. The hydrogen content was too high owing to errors in the determination of its density.

It was a very difficult experiment and a great test of Lavoisier's skill. The final solution had to wait until Morley's classical investigations just a century later.

Lavoisier lost no time in exploiting this new view of the composition of water. He saw that it explained the formation of water in organic combustions, and he was now able to use them to determine the composition of organic bodies. He determined the heat of combustion of hydrogen and saw that the oxidation of hydrogen as well as carbon in the body would account for more oxygen being consumed in respiration than the expired carbonic acid could account for. This would also explain why the amount of heat produced by an animal was greater than that calculated from its output of carbonic acid.

THE NEW CHEMISTRY AND THE REFORM OF NOMENCLATURE

The recognition of hydrogen as an element was a great advance. It explained many of Lavoisier's perplexities, and shortly afterwards Cavendish discovered that the inert part of the atmosphere, Lavoisier's 'mofette', gave nitric acid when sparked with oxygen, and Berthollet found that ammonia is a compound of 'mofette' with hydrogen.

The way was now clear. With the recognition of the elements oxygen, hydrogen, nitrogen (or azote as the French chemists called it), carbon, sulphur, phosphorus and the metals, Lavoisier could explain quite simply the composition of chemical compounds, their reactions and their quantitative relations. There were no longer obscurities to serve as a stronghold for phlogiston. Laplace and some of the younger physicists were already on his side. Of the chemists, Berthollet announced his conversion at the Academy in August 1785, followed soon by Fourcroy and Guyton de Morveau. The latter had published proposals for a reform of chemical nomenclature in 1782, but these were based on phlogiston. Discussions between the four chemists in 1786 led to their publication jointly of a new method of chemical nomenclature based on Lavoisier's theory (33). In the introduction Lavoisier acknowledges his debt to Condillac, the disciple of John Locke. From Condillac's *Logic* he had learnt the value of clarity of expression, of language as an instrument of analysis, of words as the symbols in the algebra of thought.

Till then phlogiston had been the only common basis. Names had had little reference to the nature or relationships of the substances that bore them, and many were relics of alchemy. Lavoisier's work had made a system possible, in which the names of elements like hydrogen and oxygen told something about them, and the names of compounds like sulphates and sulphites indicated their composition. The system devised by the four chemists, with few exceptions, is the one we use to-day, and its adoption helped greatly towards the general acceptance of the new theory.

TRAITÉ ÉLÉMENTAIRE

The nomenclature having been agreed upon, Lavoisier was drawn irresistibly, as he tells us, to use it to give a logical account of his new theory. His *Traité élémentaire de chimie* (34) was finished early in 1789. Its publication marks the birth of modern chemistry, and it threw a new light on all the related sciences. It was no ordinary textbook, as the first person singular appears on every page. It was in fact an autobiography of Lavoisier's work since 1772. The discoveries of seventeen years were compressed into a few pages, and the experimental basis of the new chemistry presented in such a clear compelling fashion that it won immediate acceptance. It was chemistry as Lavoisier saw it without any history or even a mention of phlogiston.

The first part outlines Lavoisier's theory of chemistry and gives the results of the crucial experiments on which it was based. 'C'est elle seule qui contient l'ensemble de la doctrine que j'ai adoptée; c'est à elle seule que j'ai cherché à donner la forme véritablement élémentaire.'

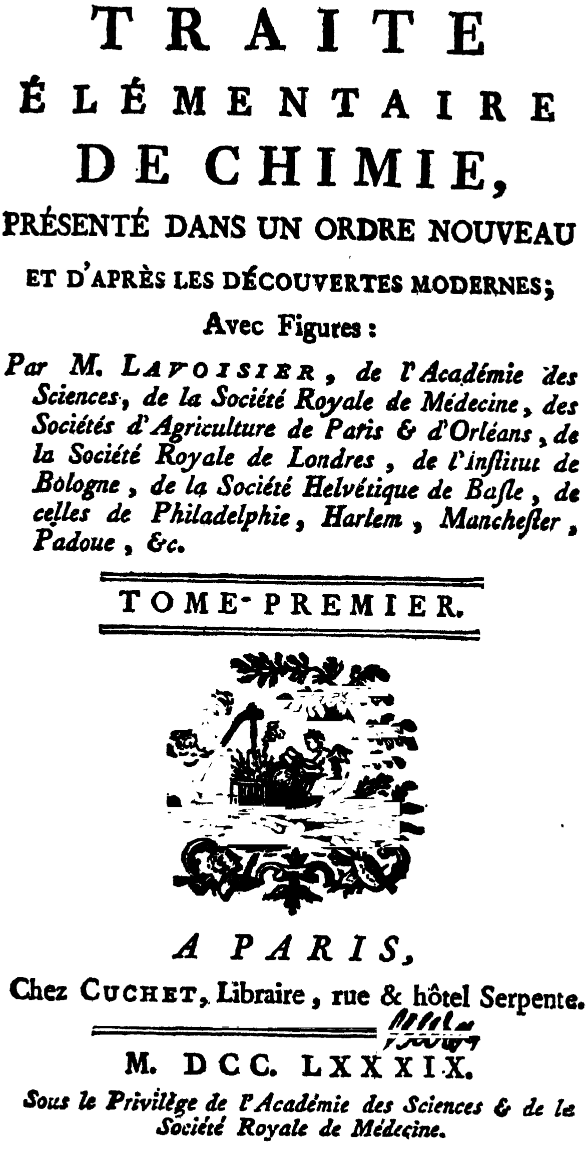
After giving his views on the nature of heat, he describes the experiments which established the composition and chemical nature of the atmosphere, water and a number of acids, gradually building up his picture on experimental results. He then sets out his views on organic chemistry, his experiments on fermentation, and his work on the analysis of organic compounds by combustion.

Oxygen is the central element, a constituent, Lavoisier thought, of all acids and all bases. Hence he predicted rightly the existence of the alkali metals isolated twenty years later by Davy. But his assumption of the presence of oxygen in all acids was less happy, and here again Davy established the facts.

The *Traité* was the forerunner of the textbooks of the nineteenth century, and some of Lavoisier's views, the dualistic nature of salts and the conception of compound radicals, were to have a profound influence on chemical thought and to be the issues on which many battles were to be fought.

The idea of the chemical equation had been implicit in all Lavoisier's quantitative work:

> Car rien ne se crée, ni dans les opérations de l'art, ni dans celles de la nature, et l'on peut poser en principe que dans toute opération, il y a une égale quantité de matière avant et après l'opération; que la qualité et la quantité des principes est la même, et qu'il n'y a que des changemens, des modifications...je puis considérer les matières mises à fermenter et le résultat obtenu après la fermentation, comme une équation algébrique.



TRAITE
ÉLÉMENTAIRE
DE CHIMIE,
PRÉSENTÉ DANS UN ORDRE NOUVEAU
ET D'APRÈS LES DÉCOUVERTES MODERNES;
Avec Figures:
Par M. Lavoisier, de l'Académie des Sciences, de la Société Royale de Médecine, des Sociétés d'Agriculture de Paris & d'Orléans, de la Société Royale de Londres, de l'Inſtitut de Bologne, de la Société Helvétique de Baſle, de celles de Philadelphie, Harlem, Mancheſter, Padoue, &c.

TOME PREMIER.

A PARIS,
Chez Cuchet, Libraire, rue & hôtel Serpente.

M. DCC. LXXXIX.
Sous le Privilège de l'Académie des Sciences & de la Société Royale de Médecine.

Figure 2. Frontispiece of the *Traité*, 1789.

The first time that Lavoisier actually used the familiar form of the chemical equation was in the chapter on alcoholic fermentation when he wrote:

Le moût de raisin = acide carbonique + alkool.

The second part of the *Traité* contains lists of salts with general methods for their preparation from each acid. It is a useful summary containing nothing 'qui me soit propre'.

The first part gave few experimental details: 'J'ai reconnu...que des descriptions minutieuses...figuroient mal dans un ouvrage de raisonnement; qu'elles interrompoient la marche des idées, et qu'elles rendoient la lecture de l'ouvrage fastidieuse et difficile.' In the third part Lavoisier gives a full account of the construction of the apparatus and the experimental methods which he had gradually developed. It is of intense interest in revealing the care and skill he devoted to his experiments, most of which were made with his own hands. Its value is enhanced by many illustrations drawn and engraved with loving care by Madame Lavoisier who had so often seen the apparatus in use in the laboratory. The book has the same individual quality as Faraday's *Chemical Manipulation*. It begins with an appeal for the use of the decimal system and ends with the prototype of future tables of physico-chemical constants.

The publication of the *Traité* marks the end of the phlogistic period and the beginning of modern chemistry. The clear logic of its presentation won immediately almost general acceptance. It was translated into several languages and became the accepted method of teaching. It had been a long fight, but in the end Lavoisier won an almost bloodless victory. In 1791 he wrote to Chaptal: 'Toute la jeunesse adopte la nouvelle théorie et j'en conclus que la révolution est faite en chimie.' Even Kirwan in 1792 wrote to Berthollet: 'Enfin, je mets bas les armes et j'abandonne la phlogistique.'

RESPIRATION AND METABOLISM

After the publication of the *Traité* Lavoisier's interests turned to organic and physiological chemistry. Almost his last scientific work was on respiration and metabolism. In 1785, in a brilliant lecture to the Society of Medicine, he described experiments on birds and guinea-pigs, showing that they could live for long periods in pure oxygen and that, to his astonishment, the rate of production of carbon dioxide was practically unaltered (35). He found later that they could live equally well in an atmosphere in which hydrogen was substituted for the nitrogen of the air. He saw that water was formed as well as carbon dioxide in the process of metabolism, thus accounting for what is now called the 'respiratory quotient', and he was able to calculate the amount of water produced. The heat evolved in its formation explained why his measurements with Laplace had shown that an animal gave out more heat than could be accounted for by the heat of formation of the carbon dioxide it expired.

From his observations on the distress of animals breathing in confined spaces as the concentration of oxygen diminished and that of carbon dioxide increased, Lavoisier's practical mind travelled to the state of the atmosphere in crowded rooms. He measured its composition in hospital wards and in the Comédie Française. Finding a deficit of oxygen and an increase of 2 or 3 % of carbon dioxide, he urged the need for better ventilation. He calculated the rate at which the atmosphere deteriorated, producing that 'impatience machinale' in an audience which was such a sad handicap to the last speaker at a meeting of the Academy.

Finally, he saw the danger that emanations from the lungs might spread disease in crowded places, and he urged the need for knowledge of the ways in which infection spreads in order to protect the health of people in large towns. The lecture was a brilliant forecast of the problems of ventilation.

Experiments on a human being were the culmination of Lavoisier's work on respiration and the subject of his last two communications at public sessions of the Academy in 1789 and 1790 (36). Seguin, his co-author, was the experimental subject. Measurements were made of his consumption of oxygen, his output of carbon dioxide, his rate of respiration and his pulse rate under different conditions. We know from sketches made by Madame Lavoisier the general experimental arrangement (plate 11). The results revealed the main factors regulating metabolism: the temperature of the environment, the output of work, and the digestion of food. Lowering the room temperature from 80 to 61° F raised the absorption of oxygen by a man at rest by 11 %, the digestion of food raised it by 50 %, doing work at the rate of 37,000 ft.lb. an hour raised it by 160 %, and the digestion of food simultaneously with work at the rate of 39,000 ft.lb. an hour raised it by 280 %. Lavoisier showed also that the increase in the pulse rate was proportional to the amount of work done, and that the amount of oxygen absorbed was proportional to the product of the pulse rate and the number of inhalations.

He saw at once the secret of the constancy of the body temperature under varying conditions of climate and occupation that had so long remained a mystery. It is regulated automatically by three balancing factors—respiration which produces heat by oxidation of carbon and hydrogen, transpiration which increases and diminishes as needed to remove any excess of heat by evaporation of water from the skin, and digestion which replaces the matter lost in the first two processes. The directness of the attack, the swift recognition of the wide implications of the results, are characteristic of Lavoisier at his best.

Finally, he discusses their bearing on the state of the body in health or disease. Fevers he thought were nature's method of restoring any disturbance of the equilibrium. The art of the doctor, he said, often consists in letting nature deal with herself, helped by suitable diet and purgatives. Contaminated atmospheres were, he thought, the cause of endemic disease and hospital and prison fevers. The best remedies were open air, freer breathing and a change of environment.

He left open the question as to whether oxidation actually took place in the lungs or in the course of circulation. Investigations already in hand on digestion and transpiration would throw further light on this.

The first paper with Seguin on transpiration (37) described an attempt to get a complete balance of the respiratory processes and ascertain the amounts of water lost through the skin and through the lungs. During the experiment Seguin was enclosed in a rubber bag, and his loss of weight was measured both with and without the bag. In addition, his absorption of oxygen and output of carbon dioxide were determined. Lavoisier could then calculate the loss of weight due to each cause on the assumption that all the oxygen absorbed formed carbon dioxide or water. If oxidation took place in the circulatory system he was uncertain whether this assumption was justified, and regarded the results as provisional.

A second paper on transpiration published after Lavoisier's death deals specially with the moisture content of air as determining comfort conditions, as to which, he said, a thermometer reading may be quite misleading (38). The paper deals also with the purpose and design of clothing.

The trend of Lavoisier's thoughts at this time are shown by the programme drafted by him for the Academy Prize of 1794 (39). His introduction deals with the problem of nutrition, the animalization, as he calls it, of vegetable and animal food. He points out the lack of knowledge of the changes that take place in the various stages of digestion, which had led the Academy to choose an investigation into the functions of the liver and bile as the subject of the Prize. A broad treatment was suggested, including the comparative anatomy of the liver and gall bladder, the chemical nature of the bile, and the pathology of the liver. The investigations were to cover chemical researches, especially the new methods of organic analysis, Lavoisier's own invention. The programme was almost a forecast of some of the major developments in physiological chemistry in the nineteenth century.

Lavoisier was the first to submit the vital functions to an exact physico-chemical analysis. He laid the foundations of physiological chemistry. His investigations into the source of animal heat first disclosed the full significance of metabolism. If he had lived on, what contributions he might still have made! But it was not to be.

POLITICS AND THE REVOLUTION

No Frenchman saw more clearly than Lavoisier the dangers that were gathering for France—the state of her finances, her outworn social structure with all its inequalities and injustice, and the selfishness of privilege with its lack of patriotism. No one had done more than Lavoisier to forecast the reforms that had to come, and when Necker in 1787 set up the Provincial Assemblies to replace the effete local Governors he saw a great opportunity. Chosen to represent the Third Estate at Orleans he was soon its outstanding figure. 'It is Lavoisier who does everything, enlivens everything, and is everywhere', said Léonce de Lavergne. We still have his memoranda—clear, practical, convincing, and amazingly modern in outlook—on the state of agriculture, the freedom of commerce, old age insurance, savings banks, infant mortality, the need for a geological survey, and above all the reform of taxation on an equitable basis (40). But privilege stood in the way and little was done.

The tide was running swiftly, and when the Estates General were summoned by the King in 1789 Lavoisier had fresh hopes. To his great disappointment he was only elected as a substitute deputy, but again he threw himself into the battle. He was a leading member of the Club of '89 which stood for constitutional monarchy. His memoir on the *Territorial Wealth of France* was ordered to be printed by the National Assembly (41). It was an attempt to estimate the national income and its taxable capacity, a pioneer effort in statistics, made possible by the information Lavoisier had collected from every province through the organization of the Ferme.

Many of the reforms Lavoisier had foreseen were soon accomplished, but early in 1790 there is already an anxious note in a letter to Benjamin Franklin: (42)

> Après vous avoir entretenu de ce qui se passe dans la chimie, ce serait bien le cas de vous parler de notre révolution politique; nous la regardons comme faitte et bien faitte sans retour;...Les personnes modérées et qui ont conservé leur sang-froid dans cette effervescence générale, pensent que les circonstances nous ont entraînés trop loin...et qu'il est à craindre que l'établissement de la nouvelle constitution n'éprouve des obstacles de la part de ceux mêmes en faveur de qui elle a été faitte....Nous regrettons bien dans ce moment votre éloignement de France; vous auriés été notre guide et vous nous auriés marqué les bornes qui nous n'aurions pas dû franchir.

In almost the last of Lavoisier's scientific papers there was a strangely prophetic passage:

> Faisons des vœux surtout pour que l'enthousiasme et l'exagération qui s'emparent si facilement des hommes réunis en assemblées nombreuses, pour que les passions humaines qui entraînent la multitude si souvent contre son propre intérêt, et qui comprennent dans leur tourbillon le sage et le philosophe commes les autres hommes, ne renversent pas un ouvrage entrepris dans si belles vues, et ne détruisent pas l'espérance de la patrie.

The violence Lavoisier feared came quickly. Members of the Club of '89 were marked men because of their moderation. He himself was attacked by Marat, whose ridiculous pamphlet, *Traité du Feu*, he had criticized. One by one he had to give up the offices which he had served so long and ably. In 1791 the administration of taxes was taken from the Ferme, and in 1792 he resigned from the Arsenal. Even the Academy was not safe. Lavoisier as its treasurer fought hard for its existence, but some of its own members turned against it, and it was suppressed in 1793.

Knowing the financial difficulties of France he offered his services in various capacities, but was refused. In 1792, however, the King wished to nominate him as Minister of the Public Funds, an office Lavoisier would have occupied so gladly in happier days. He knew it was too late, that he was suspect as a member of the old regime, and no politician. He refused, saying, 'je ne suis ni jacobin, ni feuillânt'.

Lavoisier found another outlet for his activities in these difficult years. In 1791 the National Assembly had set up a Consultative Committee on Arts and Crafts to advise the government on various questions, including useful inventions.

Lavoisier, as usual, seems to have drafted all the reports, the most important of which was entirely his own (43). It was on a system of public education, a subject to which he had given much thought, having started at his own expense a primary school at Villefrancœur.

The report dealt with the general principles of education, and its national importance. It proposed a scheme covering the whole field from the primary school to the Lycée. Under his plan all children at the age of six were to enter a primary school, and at eleven were to go on to an elementary school teaching either the arts or the arts and sciences. Finally, there were to be twelve lycées giving the highest form of public education with a wide choice of subjects. Lavoisier had an eye to practical subjects like hygiene, the weather and simple surveying. Country schools were to teach the elements of agriculture and town schools the elements of commerce.

Lavoisier is emphatic as to the importance of education in every walk of life.

> Le mot *industrie* n'exprime pas toujours un emploi de forces, ni même d'adresse, il exprime le plus souvent un emploi des facultés de l'esprit....Le cultivateur qui prospère le plus n'est pas toujours celui qui est physiquement le plus fort et le plus adroit; c'est celui qui est le plus intelligent....Organisez l'instruction publique dans toutes ses parties; donnez du mouvement aux arts, aux sciences, à l'industrie, au commerce.

In 1791 the National Assembly had entrusted to the Academy the task of establishing a uniform system of weights and measures, a very necessary reform which resulted in the metric system. A commission was set up consisting of Borda, Lagrange, Laplace, Monge and Condorcet with Lavoisier as secretary and treasurer. The work was divided between various groups; Lavoisier and Haüy were to determine the density of distilled water at zero temperature. Lavoisier was also responsible for the administration of the commission, which continued after the suppression of the Academy. It was his last scientific task (44).

Meanwhile, in the general attack on the *ancien régime*, feeling was running high against the members of the Ferme, and in 1793 they were arrested on charges of maladministration. Lavoisier asked to be allowed to continue his work on the standards, and Haüy and Borda pleaded for him in vain to the Convention, risking their own lives. All his wife's efforts, too, were unsuccessful.

When the charges against the Ferme were formulated, it was Lavoisier who drew up their defence (45). He gave a masterly review of their financial transactions, showing that they had done their work efficiently and without undue profit to themselves. But it was the height of the Terror, the trial was a mockery of justice, and Lavoisier and twenty-seven of his colleagues were guillotined a few hours later. Lavoisier followed his father-in-law, Paulze, to the scaffold, and met death with the courage and philosophy of his generation.

Thus died this great Frenchman in his fifty-first year at the height of his achievements. The loss to France and to the world was immeasurable. The Terror had nearly run its course; within a few years came reconstruction, when under Napoleon science once more came into its own. What a contribution Lavoisier might have made to the Industrial Revolution in France, with his vision of her needs, his imaginative use of science in industry, agriculture and hygiene, and his administrative skill and experience.

If in the fields of politics and economics he was cut off before his work came to fruition, he had foreseen the changes soon to come in industry, in social legislation and in education. His eloquence and writings must have left their mark on the minds of those who were to remake France.

In science it was otherwise. There his work was done with almost incredible effectiveness and speed. No other revolution in scientific thought has ever been so complete, so swift, or has done so much to clear men's minds of the cobwebs of centuries. He quickened the advance of science over its whole front. Not only chemists, but workers in physics, biology, medicine and agriculture were given for the first time a clear view of the various forms of matter we call elements, and of the distinction between chemical and physical changes. Lavoisier put new and potent weapons into their hands, just when they were needed for science to play its full part in the nineteenth century. Compare the clear-cut vision of Lavoisier's great treatise, based on quantitative measurement, on the methods of analysis and synthesis, and on the notion of the chemical equation, with the confused picture of chemistry in Macquer's textbook, with no general conception but the theory of phlogiston. It was no wonder that progress was quickened. Dalton's atomic theory, Davy's discovery of the alkali metals, his recognition of the halogens as elements, and the massive contributions of Berzelius followed swiftly, completing the picture and giving chemistry its modern form.

It was, indeed, a revolution, and I have tried to show how Lavoisier fought the battle for the New Chemistry almost single-handed, for he had no school, with many other preoccupations in one of the busiest of lives. He was fortunate in his great contemporaries, Black, Priestley and Cavendish, who so often gave him the clue he was searching for, but he alone saw the significance of their discoveries.

What was the secret of his greatness? Lavoisier was above all things a reformer with an intense desire to improve everything he saw around him, whether it was the confused state of chemistry, the system of taxation, the lot of the peasant, the making of gunpowder or the methods of agriculture. He had the logical systematic mind of his countrymen, joined with great constructive power and a creative urge. Burning curiosity made him a born researcher. Every problem was to him a challenge for research and experiment, and for him research meant measurement, the method of modern science. He had an essentially modern outlook. He was not bound by tradition or authority. He looked his problems squarely in the face, and judged them on evidence with the shrewd common sense he showed as a man of affairs.

Lavoisier's great strength was his realism. He saw each problem in its smallest details; and no one examined them with greater care, for he was a great experimenter; but he also had great vision and saw his problems in their widest implications. He gave his visions quantitative form, in chemistry with the balance, in finance with statistics. He applied to any practical problem the same analysis and measurement that he used in the laboratory. It is often said that scientists are too remote from real life. Lavoisier is the outstanding example of the value of the laboratory mind in action. That is his lesson for us to-day.

I am greatly indebted to previous biographers of Lavoisier, and especially to the late Professor A. N. Meldrum for his scholarly researches on Lavoisier's early work. I have to thank also my secretary, Miss Josephine Wasse, without whose never-failing help the preparation of this lecture during wartime would have been impossible.

Notes

(1) *Œuvres de Lavoisier*, 1864–93, **4**, 1. Paris: Imprimerie Impériale.
(2) *Œuvres*, **3**, 765.
(3) *Œuvres*, **5**, 187.
(4) *Œuvres*, **3**, 1.
(5) *Œuvres*, **3** and **4**.
(6) Arthur Young, *Travels in France and Italy*, 1934, p. 78. London: Dent.
(7) Gouverneur Morris, *A Diary of the French Revolution*, 1939, *passim*. London: Harrap.
(8) *Œuvres*, **3**, 106, 128.
(9) *Œuvres*, **2**, 1.
(10) *Œuvres*, **2**, 38.
(11) *Œuvres*, **2**, 64.
(12) A. N. Meldrum, *The Eighteenth Century Revolution in Science—the First Phase*, 1930. Calcutta: Longmans, Green. Lavoisier's three Notes on Combustion: 1772. *Archeion*, 1932, **14**, 16.
(13) A. N. Meldrum, op. cit.
(14) *Œuvres*, **2**, 103.
(15) *Œuvres*, **1**, 439.
(16) Berthelot, *La Révolution Chimique: Lavoisier*, 1890, p. 48. Paris: Germer Baillière.
(17) Berthelot, op. cit. p. 246.
(18) Robert Boyle, *Essays of the Strange Subtility*
Determinate Nature } *of Effluviums*
Great Efficacy
To which are annext New Experiments to make fire and flame ponderable: together with a Discovery of the perviousness of glass, 1673, *passim*. London: W. G. for M. Pitt.
(19) Lavoisier, *Observations sur la Physique*, edited Abbé Rozier, 1774, **4**, 448.
(20) *Œuvres*, **2**, 105.
(21) Berthelot, op. cit. p. 257.
(22) *Œuvres*, **2**, 122.
(23) Joseph Priestley, *Experiments and Observations on Different Kinds of Air*, 1774, **2**, 320. London: J. Johnson.
(24) *Œuvres*, **5**, 391.
(25) *Œuvres*, **2**, 129.
(26) *Œuvres*, **2**, 812.
(27) Lavoisier, *Observations sur la Physique*, 1772, **2**, 510.
(28) Lavoisier, *Observations sur la Physique*, 1773, **1**, 10.
(29) *Œuvres*, **2**, 283.
(30) *Œuvres*, **2**, 334.
(31) Arthur Young, op. cit. p. 78.
(32) *Journal Polytype*, 5 February 1786.

(33) Guyton de Morveau, Lavoisier, Berthollet and de Fourcroy, *Méthode de Nomenclature Chimique*, 1787. Paris.
(34) *Œuvres*, **1**, 1.
(35) *Œuvres*, **2**, 676.
(36) *Œuvres*, **2**, 688; *Annales de Chimie*, 1814, **91**, 318.
(37) *Œuvres*, **2**, 704.
(38) Lavoisier, *Annales de Chimie*, 1814, **90**, 5.
(39) *Œuvres*, **6**, 33.
(40) *Œuvres*, **6**, 238.
(41) *Œuvres*, **6**, 403.
(42) E. Grimaux, *Lavoisier*, 1899, p. 201. Paris: Germer Baillière.
(43) *Œuvres*, **6**, 516.
(44) *Œuvres*, **6**, 670.
(45) *Œuvres*, **6**, 570.

Determination of the velocity coefficients for polymerization processes.

I. The direct photopolymerization of vinyl acetate

By G. M. Burnett and H. W. Melville, F.R.S.
Chemistry Department, University of Aberdeen

(*Received* 6 *November* 1946—*Read* 6 *February* 1947)

This paper describes a method whereby the individual velocity coefficients of a vinyl polymerization reaction have been determined by a detailed analysis of the kinetics of the photo-polymerization of liquid vinyl acetate. The evaluation of the coefficients is made by the development of an accurate method for measuring the life time of the active polymeric molecules employing intermittent radiation. In addition, the kinetic chain length and the number average molecular weight of the polymers have been measured. It is shown that each absorbed quantum of radiation activates one vinyl acetate molecule probably as a diradical. The chain terminates by a disproportionation reaction between the polymer radicals. The size of the growth or propagation bimolecular coefficient decreases only slightly as the molecule grows, the energy of activation is 4400 cal. and the steric factor of the growth reaction is 10^{-5}. In the termination reaction there is likewise only a small decrease of velocity coefficient with increasing molecular size—no energy of activation is required and the steric factor has the surprisingly large value of 10^{-2}.

INTRODUCTION

The three fundamental reactions controlling the velocity of a chain polymerization process are: (*a*) initiation reaction, (*b*) growth reaction, and (*c*) termination of growth. Hitherto the kinetic analysis has made use of two measurable quantities, namely, (i) the overall rate of reaction, and (ii) the rate of initiation. The ratio of (i) to (ii) will give the kinetic chain length which is, in some cases, equal to the degree of polymerization of the polymer. There are, however, three unknowns involved in the analysis. Until a third parameter can be measured, further analysis is impossible, and, in particular, the values of the individual growth and termination coefficients cannot be obtained. The third parameter which requires to be measured is the stationary concentration of the active polymer. If this problem is soluble it is practicable to calculate in moles, litres and seconds the value of the growth coefficient and thereby obtain a numerical measure of the tendency of a molecule to polymerize, according to a given mechanism. Then it should be possible to see much more clearly how chemical constitution affects reactivity. Similar information automatically becomes available regarding the termination coefficient, thereby providing further data about the ultimate fate of the highly active polymer molecules. Precise measurements of the kinetic chain length, i.e. the number of monomeric units polymerized for each molecule originally activated, and the degree of polymerization will enable transfer reaction to be detected, for in such a case the chain length will be greater than the degree of polymerization.

By standard technique the value of the chain length can easily be varied, and therefore it may be possible to find whether the growth and termination coefficients change with the size of the molecule. This information makes it possible to compute accurately what the distribution of the molecular sizes will be in the resultant polymer.

An outline of how this problem can be approached has already been given (Jones & Melville 1940). The present paper describes an accurate method by which the above-mentioned problems may be quantitatively solved for the case of the polymerization of pure liquid vinyl acetate.

EXPERIMENTAL

In many polymerization reactions it has been established that the presence of oxygen has a considerable effect on the course of the reaction, the effect depending to some extent on the length of time in which the monomer is in contact with the oxygen. In view of this observation all the experiments were carried out *in vacuo* and all the materials were carefully freed from air before use. Evacuation of the apparatus was effected by means of a mercury diffusion pump backed by a rotary oil pump, with the customary liquid-air trap and drying tube in position. Reservoirs of the monomer and various solvents were sealed to the main vacuum line, and taps leading to the reaction system were covered by mercury cut-offs. In certain cases, silicone tap grease, thickened with talc, was used on the taps, otherwise Apiezon L was employed.

The reaction vessel was made of fused silica and was of the design shown in figure 1. The ends of the vessel were optically flat to enable accurate measurement of the quantum input to be made. The vessel was calibrated so as to act as a dilatometer. For reactions at high temperatures and involving solvents sealed tubes were used and were sealed off *in vacuo*. Thermostatic control was achieved by surrounding the reaction vessel with a copper bath, fitted with silica windows, through which water from a thermostat could be circulated.

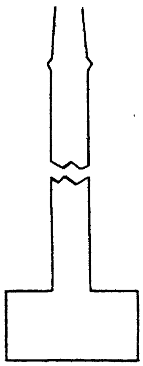


FIGURE 1. The reaction vessel.

The lamp used throughout the experiments was a G.E.C. 'Osira', from which the outer glass envelope was removed. The lamp then consisted of a high-pressure mercury arc which ran at 90 V and 1·3 amp. Fluctuations in the output were avoided by feeding the lamp from a constant-voltage transformer, and by efficient shielding from draughts. The shield was made so as to act as a parabolic reflector also, giving, therefore, an almost parallel beam of light which was focused on the reaction vessel by means of a silica lens of short focal length. The polymerization reaction depends to a great extent on the wave-length of the light employed. Thus at very short wave-lengths there appears to be very considerable decomposition of the monomer which seems to result in the production of either a retarder or inhibitor of the polymerization process. To avoid this as far as possible a filter of 50 % acetic acid in water was inserted in the path of the light, thus effectively removing all light of low wave-length.

Monomeric vinyl acetate of commercial grade was obtained from Shawanigan Ltd. Hydrolysis of the monomer results in the formation of acetaldehyde which has been shown to be an inhibitor of the gas-phase reaction so that its complete removal is essential. Purification was carried out by the fractional distillation of the crude monomer until the distillate gave only a very faint coloration with Schiff's reagent on standing for 5 min. Only the cut between 72 and 73° C was retained. This was redistilled at atmospheric pressure and stored in a tightly stoppered dark bottle. A suitable quantity of the purified monomer was transferred to the reservoir on the vacuum apparatus and sealed off. The monomer was frozen in liquid air and the reservoir evacuated, then isolated from the pumps and the liquid air removed. The dissolved air escaped and the process was repeated until no trace of air was noted. Finally, the first 5 c.c. or so of the monomer was distilled directly to the liquid-air trap further to ensure purity. Small quantities of monomer received from Boake Roberts Ltd. and I.C.I. Ltd. were similarly treated and gave the same experimental results, thereby establishing the method of purification as being adequate.

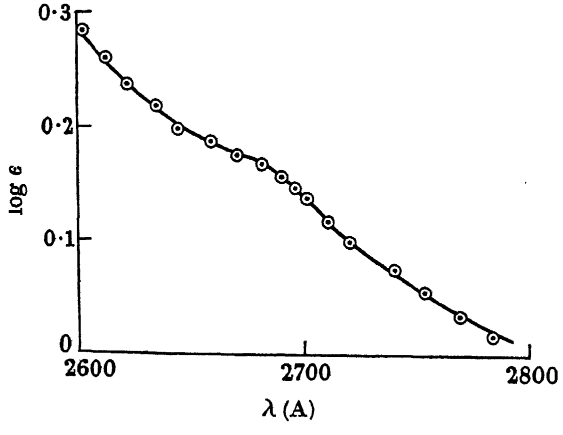


Figure 2. The ultra-violet absorption spectrum of vinyl acetate.
ϵ = Molar decadic extinction coefficient.

Benzoquinone which was used as an inhibitor was purified by steam distillation, followed by thorough drying and sublimation. For convenience it was made up temporarily in freshly distilled acetone.

The ultra-violet absorption spectrum of the monomer

The absorption spectrum was photographed on a Hilger quartz spectrograph using *n*-hexane of spectroscopic purity as the comparison liquid. The solvent and solution were contained in microcells, 5 mm. in length. The usual technique was followed in mapping the absorption spectrum which is shown in figure 2. It will be seen that the value of the extinction coefficient falls fairly rapidly at first and then more slowly. This is somewhat the same as in the case of methyl methacrylate (Melville 1937) and was then associated with the double bond and the $-COO.CH_3$ group present in that molecule. Similar results with other molecules which do not contain an ester group suggests, however, that the two parts of the curve are the characteristic of the molecule rather than a part of it.

The bulk photopolymerization of vinyl acetate

Since the density of the polymer formed is independent of the molecular weight of the macromolecule, it is convenient to use the density change as an accurate measure of the course of the reaction. This is most readily accomplished by measuring the contraction of the monomer as it polymerizes, the total volume contraction being 26·82 % for 100 % polymerization at 25° C (Starkweather & Taylor 1930). The reaction vessel was of such dimensions that a drop of 0·01 mm. which could be detected by the cathetometer used, corresponded to an extent of polymerization of about 0·03 %. Checks of the final amount of polymer were made by the determination of either the refractive index of the mixture, using the results of Breitenbach & Raff (1939), or by the evaporation of the monomer under high vacuum from the mixture of the monomer and polymer, and weighing the polymer remaining.

The course of the reaction over the initial stages is shown in figure 3. It will be seen from table 1 that the reaction is of zero order with respect to the monomer concentration. It was decided that it was not worth carrying the reaction further than this stage, since the increasing viscosity of the solution made the meniscus shape change considerably, and in the viscous liquid temperature control became much more difficult because of the hindrance of the convection currents and the poor thermal conductivity of the reaction mixture. Also as the conversion becomes greater there is the possibility of an explosion like polymerization (Schulz 1941), and that this does occur is shown in figure 4. The zero order of the reaction over such a large percentage of reaction is, at first sight, surprising, but the explanation would appear to be as follows. It will be shown in a later paper that the rate of reaction is proportional to the first power of the monomer concentration (naturally this cannot be tested with the pure monomer, since the value of the monomer concentration is fixed by the density of the liquid). During polymerization the monomer concentration remains practically unaltered, for the gradual accumulation of

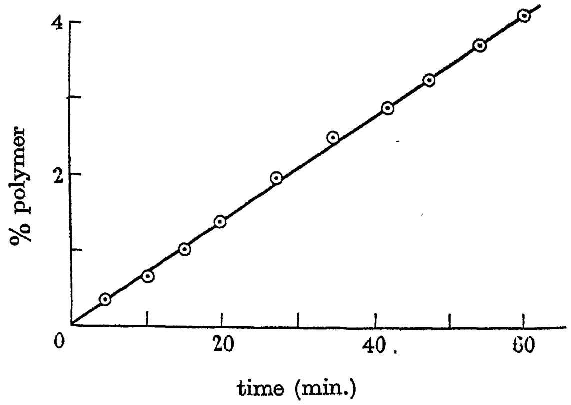


FIGURE 3. The direct reaction (initial stages).

TABLE 1. THE COURSE OF POLYMERIZATION

Intensity of illumination = 5×10^{15} $h\nu$/sec. Temperature = 15·9° C.

time (min.) (*T*)	contraction %	conversion % (*C*)	*C*/*T*
0	—	—	—
4	0·070	0·26	0·065
8	0·142	0·53	0·066
12	0·218	0·81	0·067
20	0·360	1·34	0·067
25	0·440	1·64	0·066
35	0·617	2·30	0·066
45	0·805	3·00	0·067
60	1·080	4·01	0·067

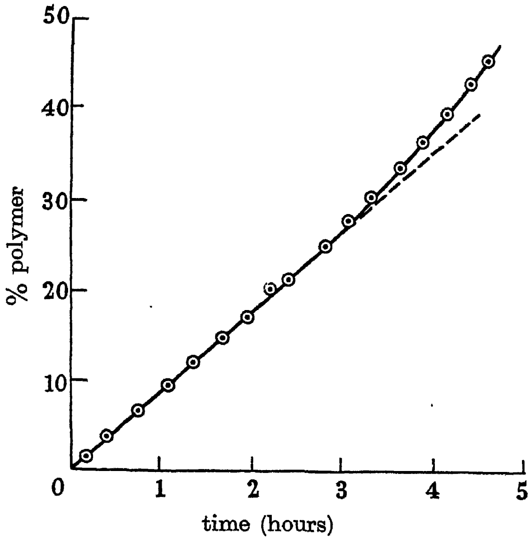


FIGURE 4. The direct reaction showing the 'gel effect'.

polymer causes no significant reduction in the vapour pressure of this mixture owing to its high molecular weight. Similar results have been obtained in the catalyzed polymerization of methyl methacrylate (Norrish & Brookman 1937).

If the absorption of the activating radiation is dependent on the concentration of monomer, the results shown in table 2 may be readily obtained. This shows definitely that it is not possible with certainty to derive information on the nature of the termination reaction without investigating the overall kinetics to a much greater extent.

TABLE 2

termination reaction	lowest order of reaction (absorption independent of monomer conc.)	highest order of reaction (absorption $\alpha(M)$)
spontaneous	1	2
monomer	0	1
mutual	1	1·5

The intensity exponent of the polymerization will give a definite clue to the nature of the chain-breaking process. For either of the first two cases in table 2 the rate of reaction is proportional to the light intensity, while for the third case the rate is proportional to the square root of the intensity of absorbed radiation. Hence a distinction can be readily drawn.

The intensity of the light was varied by means of perforated screens whose transmission was measured by means of a photocell. The distance apart of the holes was such that no interference could be caused due to distinct beams of light, as has been described (Jones & Melville 1940). In each case the rates under the total radiation and with the screen in position were measured, and in figure 5 the log relative value of the rate is plotted as a function of the log relative value of the intensity. It is easy to show that the relationship ought to be a straight line whose slope gives the value of the intensity exponent, since

$$R = k\{f(I)\}^n,$$

where R is the rate, k a proportionality constant, $f(I)$ some function of the light intensity, and n the intensity exponent. Putting the above expression in logarithmic form gives

$$\log R = \log k + n \log f(I),$$

so that the log plot must obviously be a linear one as described above. From the graph of figure 5 the value of the intensity exponent works out at 0·50.

Using the kinetic theory of Melville & Gee (1944) this implies that the termination reaction involves the mutual destruction of two of the growing polymer molecules. Assuming that the chain carrier is a free radical (evidence for which will be given in a subsequent paper), it is obvious that termination may take place by two distinct mechanisms. First, the radicals may simply combine to give one inert or 'dead' polymer, by the scheme

$$\text{—CH}_2\text{—CH}X\text{—} + \text{—}X\text{HC—CH}_2\text{—} \rightarrow \text{—CH}_2.\text{CH}X.X\text{HC}.\text{CH}_2\text{—},$$

or, secondly, the reaction may be such that one of the free radicals dehydrogenates the other so that both molecules are stabilized, thus

$$—CH_2.CHX— + —XHC.CH_2— \rightarrow —CH_2.CH_2X + CHX{=}CH—.$$

If, as is most likely in this case, the initial step is the formation of a free diradical, i.e.

$$CH_2.CHX + h\nu = —CH_2.CHX—,$$

then it is obvious that the most likely—if not the only—mechanism for mutual termination is the second of the above. Unfortunately, the value of the intensity exponent does not lead to any conclusive mechanism, since for both the above mechanisms the value of the exponent is 0·5.

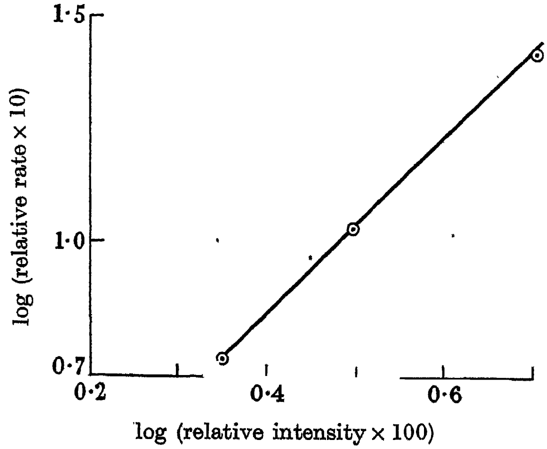


FIGURE 5. The intensity exponent.

It is interesting that in this case the value is so close to 0·5, since it excludes the possibility of any other type of termination mechanism operating simultaneously. In the case of the vapour phase polymerization of vinyl acetate the value of the intensity exponent is appreciably greater than 0·5, which would suggest that there is a considerable amount of termination by some mechanism other than mutual—probably by the walls of the vessel. For the catalyzed polymerization of vinyl acetate again it is found that the rate of reaction is accurately proportional to the square root of the catalyst concentration, again implying that a mutual termination mechanism is solely responsible for the breaking of chains (Cuthbertson, Gee & Rideal 1937).

The quantum yield

As a quantum counter in these experiments a Baird television photocell, fitted with a silica envelope, was used. In the set-up employed the photocurrent flowed through a number of resistances in series in order to facilitate the use of the instrument over a wide range of light intensities. The potential difference over a suitable resistance was measured on a Marconi type valve voltmeter, capable of reading to

0·1 mV. The high-tension supply for the photocell was an ordinary 120 V dry battery, the voltage of which was frequently checked in order to ensure constant running conditions for the photocell. All the comparative measurements were made on the linear portion of the characteristic curve for the photocell.

The photocell was calibrated for various lines in the mercury spectrum against uranyl oxalate, using the following technique. The cell was mounted in a brass tube which was fitted with an aperture about 7·5 mm. in diameter. Light from a high-pressure arc was focused on the aperture by means of a pair of silica lenses in such a way that the diameter of the beam was just equal to that of the opening, and the image fell centrally on the sensitive plate of the photocell. Using a suitable filter, a given line or narrow band in the ultra-violet region could be isolated. After a reading had been obtained on the voltmeter, the photocell was removed and replaced by a shallow dish containing a known amount of uranyl oxalate solution. After a definite period of illumination—usually about 6 hr.—the oxalate was titrated against a standard solution of potassium permanganate at 80° C. Comparison with a blank titration made the computation of the light intensity possible with reference to the quantum yields of the oxalate reaction. It was found that the sensitivity of the cell increased rapidly at first and then less rapidly and finally reached a maximum in the visible blue region.

Losses of light in these experiments could result from (*a*) reflexion at the various interfaces and (*b*) absorption by the water of the thermostat bath. Losses by reflexion were found to be negligible by comparison of the light transmitted through the empty reaction vessel and with the vessel filled with *n*-hexane of spectroscopic purity. This is not altogether surprising, since the refractive indices of the liquid and silica are not very much different. Losses due to absorption by the water amounted to some 7 % of the total radiation, by a similar comparative measurement.

A silica lens was used to bring the light into a slightly converging beam of approximately the same diameter as the reaction vessel. The final image again fell centrally on the sensitive plate of the photocell. The light beam required to be slightly converging in order to fulfil the final condition. All light which did not pass through the reaction system was prevented from reaching the photocell by means of a light metal shield, with a sleeve, which fitted round the end of the reaction vessel. The intensity of the ultra-violet light passing through the cell was measured both with the cell empty and filled with monomer, from which the intensity of the absorbed radiation could be found. Frequent checks were made in order to ensure that the running of the lamp was perfectly steady and the light absorbed was always constant.

The quantity of polymer produced was determined by measurement of the refractive index of the mixture on a Pulfrich refractometer, using the sodium D line as a source of illumination. Table 3 gives the results obtained from these experiments.

The figure in the bracket in column 3 of the table was obtained by evaporating off the monomer under high vacuum and weighing the residual polymer. It will be noted from this table that the values of the kinetic chain length, ν_k, are in the ratio of 6·6:1, while the ratio of the square roots of the corresponding intensities is 7:1, which is in good agreement with the theory.

TABLE 3. THE QUANTUM YIELD

Temperature = 15·9° C. $\lambda = 2800$ to 3200 A.

intensity of illumination ($h\nu$/sec.)	time of illumination (sec.)	wt. of polymer (g.)	ν_k
$9{\cdot}83 \times 10^{13}$	24,800	0·456	1500
	39,500	0·714	1460
	11,500	0·227 (0·222)	1440 (1460)
$4{\cdot}90 \times 10^{15}$	3,600	0·476	220
	1,500	0·204	228
	3,600	0·481	222 (223)

Assuming that no transfer takes place in this reaction, it should be possible to obtain a value for the primary quantum efficiency from a determination of the number average molecular weight of the final product. This is, however, complicated by the fact that the nature of the termination reaction is not known with certainty. If the breaking of the chains should result in the formation of two molecules of 'dead' polymer, then this method will give the true value of the primary quantum efficiency. If, on the other hand, the termination reaction should bring about the formation of only one inactive molecule due to the combination of two of the growing chains, then the molecular weight would obviously be, on the average, twice the kinetic molecular weight, so that the value of the quantum efficiency would be half of the true value. Osmometric measurements on the samples obtained from the first three experiments are quoted in table 3. The osmometer used was of the type suggested by Fuoss & Mead (1943) and had a vertical capillary. The membrane used was bacterial cellulose, details of whose production, treatment and use have already been given (Cruickshank, Masson, Melville & Menzies 1946). The results obtained by this method give a molecular weight of 120,000 ± 4000 or a chain length of 1400 ± 50, which is in comparatively good agreement with the kinetic chain length and suggests that the primary quantum efficiency is probably unity. This means that the number of chains started per second is equal to the quantum input per second.

In order to verify this conclusion as completely as possible it is necessary to devise a means of ensuring that the number of chains started can be counted. By the choice of a sufficiently strong inhibitor this becomes possible. It has long been known that *p*-benzoquinone acts as an inhibitor towards the thermal polymerization of styrene (Foord 1940). Since it is highly probable that the chain carrier in this case and also in the photopolymerization of vinyl acetate are of the same type, *p*-benzoquinone was tried as an inhibitor of this reaction. The ultra-violet absorption spectrum of the quinone has a minimum in the region under discussion, so that there can be no internal filter effect (Marchlewski & Moroz 1924). The following general observations were made on the behaviour of the quinone in this reaction. Inhibition is complete over a considerable period whose duration depends on the concentration of the quinone initially present, and during this period the colour of the solution fades and finally disappears. On the disappearance of the colour the rate of polymerization is the same as that which is found without the inhibitor. This means that it is impossible for the inhibition to result in the formation of a compound which may act as an inhibitor or even retarder of the polymerization reaction. Table 4 gives the results for the experiments, carried out using a light intensity of 5·0 × 10^{15} $h\nu$/sec. at 25° C.

TABLE 4. INHIBITION BY QUINONE

Weight of monomer = 11·18 g.

wt. of quinone (mg.) (*W*)	duration of inhibition (min.) (*T*)	*T*/*W*
2·4	48	20·0
5·0	103	20·6
7·5	148	19·9

In figure 6 is plotted the course of a number of reactions which shows clearly the effect of the addition of the quinone. The curves obtained by Foord for the thermal polymerization of styrene follow a somewhat similar course for benzoquinone inhibition. Foord found that other quinones were not so effective, but, unfortunately, this cannot be tested here because of the absorption spectra which contain maxima in the region of 3000 A in the case of the other members of the quinone family.

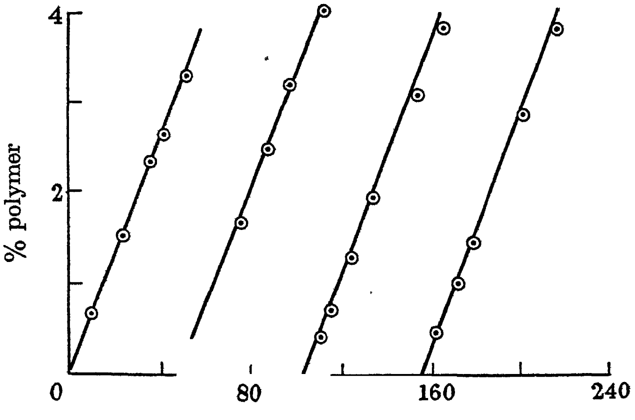


FIGURE 6. Inhibition by *p*-benzoquinone.

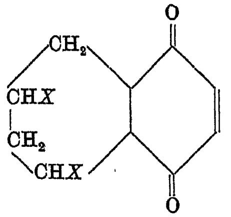
Experiments of the same type were carried out under the conditions employed for the quantum yield experiments. The same proportionality was observed, and the results of the experiments are shown in table 5.

In the calculation of the primary quantum efficiency in this table it has been tacitly assumed that only one molecule of the quinone is required for the breaking of the chain. The stoichiometric analysis of the reaction by Goldfinger *et al.* (1943) tends to justify this assumption. Also it is evident from the shape of the curves in figure 6 that the initial growth before stopping can be beyond probably a dimer, but even if a dimer is formed this has no effect on the calculations given above. Chemically the reaction still requires further investigation, although some points may be made here in order to clarify the situation.

TABLE 5. PRIMARY QUANTUM EFFICIENCY FROM QUINONE INHIBITION

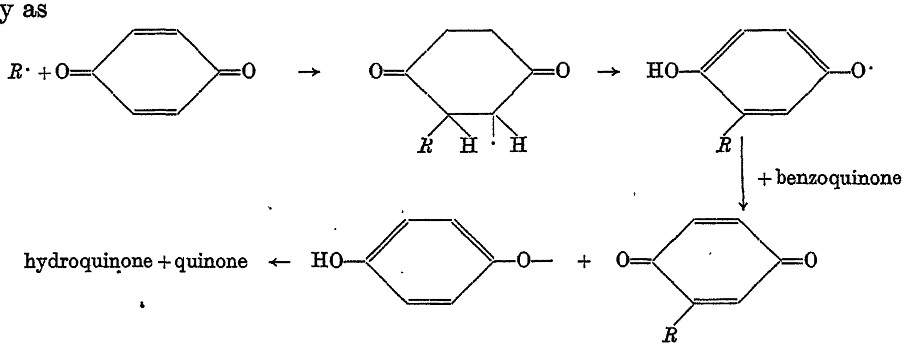
wt. of quinone (mg.)	duration of inhibition (min.)	primary efficiency
0·75	670	1·00
0·13	112	0·98
0·29	250	0·98

The mode of action of quinone appears to be twofold in the inhibition of polymerization reactions. If the reaction is initiated thermally or photochemically then the results which have been given here are applicable. If, however, initiation is by a catalyst, such as a peroxide, the final state is reached only slowly. There is initially a period of complete inhibition followed by polymerization, and the rate of polymerization increases slowly towards a maximum. We suggest that the difference is due to a difference in the nature of the chain carrier. In the first case it is almost certainly a diradical and in the second a monoradical. The dimeric diradical could undergo a modified Diels-Alder reaction with the quinone, thus:



Now it is known that the addition of a further such diradical is unlikely, so that the fact that inhibition or retardation is not caused by the product is accounted for. In the case of styrene some progress was made along these lines, and a compound of molecular weight 310 and constitution $C_{22}H_{18}O_2$ was isolated. This conforms to the above if $X = C_6H_5$. It is interesting to note that in the case of the catalyzed reaction considerable quantities of hydroquinone were isolated. This is an efficient retarder of the photopolymerization of vinyl acetate (Vernon & Taylor 1931), so that further support is given to the ideas which have just been put forward.

The normal reaction between a free monoradical and a quinone molecule will account for the formation of hydroquinone. This reaction may be shown schematically as



but this appears to have some weakness in that a substituted quinone is formed which has probably a high colour—which is not supported experimentally—and more than one quinone molecule is used to stop the chain.

The energy of activation

In gas-phase polymerization processes the measured energy of activation always appears to be negative, which is not in keeping with the general kinetic conceptions. This may be due to the fact that in such reactions the formation of the polymer implies the formation of a new phase upon which adsorption of the monomer may take place. As the heat of adsorption is, in most cases, unknown, no correction can be applied. In a system which remains homogeneous during the course of the reaction, it would be expected that normal values of the energy of activation would be obtained.

In this case the rate of the reaction was measured in sealed tubes at temperatures from 18 to 45° C. Corrections were applied to allow for the density changes at the different temperatures although these were not very great. The reaction velocity was found to increase as the temperature increased, thus giving rise to a positive value for the energy of activation. The usual plot of log rate against the reciprocal of the absolute temperature gives a value of 4·40 kcal./g.mol. for the energy of activation. The results are set out in detail in table 6.

Table 6. The energy of activation

temperature °K	relative rate	log (relative rate)	*E* (kcal./g.mol.)
318	2·02	0·305	4·21
310	1·71	0·232	4·48
302	1·42	0·154	4·40
295	1·19	0·078	4·48
288	1·00	0·000	

The value for the energy of activation agrees with that obtained by Vernon & Taylor (1931), who found a value of approximately 4 kcal./g.mol. The positive value of the activation energy appears to support the view that for gas-phase reactions a correction due to the heat of adsorption ought to be made. This had been done in the case of the polymerization of chloroprene (Bolland & Melville 1938) when the corrected value is of the same order as that given here. It will be shown in a later section that this energy of activation is wholly due to the propagation step of the reaction, and since the primary quantum efficiency is unity, i.e. as high as possible, it is probable that the rate of initiation is temperature independent, so that the energy of activation of the termination process is zero.

Transfer reactions

Although it has been stated in several papers that transfer is liable to occur in bulk polymerization, there is little in the way of concrete evidence to support the view. Although it has been shown that there is no detectable transfer in this reaction it might be of interest to work out the formal kinetics of this type of reaction on the lines adopted by Melville & Gee (1944).

The reaction steps may be formulated as

$$\begin{array}{lll} M+h\nu = P_1 & \text{initiation} & f(I) \\ \left.\begin{array}{l} P_1+M = P_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ P_n+M = P_{n+1} \end{array}\right\} & \text{propagation} & \begin{array}{l} k_{p_1} \\ \\ k_{p_n} \end{array} \\ P_n+M = M_n+P_1 & \text{transfer} & k_{f_n} \\ P_n+P_m = M_m+M_n & \text{termination} & k_{m_n}, k_{m_m} \end{array}$$

where the symbols have their usual significance. Since the reaction ceases when the light is removed, it follows that the stationary state method of analysis can be used. Therefore

$$\frac{d(P_1)}{dt} = f(I) + (M)\,\Sigma k_{f_s}(P_s) - k_{p_1}(P_1)\,(M) - k_{f_1}(P_1)\,(M) - k_{m_1}(P_1)\,\Sigma k_{m_s}(P_s) = 0,$$

or, generally,

$$\frac{d(P_n)}{dt} = k_{p_{n-1}}(P_{n-1})\,(M) - k_{p_n}(P_n)\,(M) - k_{f_n}(P_n)\,(M) - k_{m_n}(P_n)\,\Sigma k_{m_s}(P_s) = 0.$$

By adding these equations to infinity then

$$f(I) = \{\Sigma k_{m_s}(P_s)\}^2,$$

or

$$\{f(I)\}^{\frac{1}{2}} = \Sigma k_{m_s}(P_s).$$

The rate of reaction will be given by

$$-\frac{d(M)}{dt} = f(I) + (M)\,\Sigma k_{p_s}(P_s) + (M)\,\Sigma k_{f_s}(P_s).$$

Now, assuming, after the manner of Melville & Gee, that

$$k_{p_n} = \delta^{-1} k_{m_n} \quad \text{and} \quad k_{f_n} = \mu^{-1} k_{m_n},$$

where δ and μ are both independent of the molecular size, then

$$-\frac{d(M)}{dt} = f(I) + (M)\,\delta^{-1}\Sigma k_{m_s}(P_s) + (M)\,\mu^{-1}\Sigma k_{m_s}(P_s).$$

If, as is general, the chains are very long, then the first term may be omitted and substitution of the value of $\Sigma k_{m_s}(P)_s$ found above will lead to

$$-\frac{d(M)}{dt} = (M)\{f(I)\}^{\frac{1}{2}}\{\mu^{-1}+\delta^{-1}\}.$$

This means that the reaction will still be proportional to the square root of the light intensity providing that the mutual termination mechanism is still operative for the final breaking of the chains.

DETERMINATION OF THE LIFE TIME OF THE ACTIVE POLYMER

As was pointed out earlier in this paper, the overall rate for the bulk photopolymerization of vinyl acetate is proportional to the square root of the intensity of the absorbed radiation. This feature of the reaction makes it possible to obtain experimentally the value of the life time of the growing chain. The fact that the reaction velocity is constant during the initial stages of the reaction makes this a comparatively straightforward matter. The use of intermittent illumination to achieve the measurement of the life time of the active particle was suggested by Chapman, Briers & Walters (1926), but his mathematical analysis made it necessary to rely on trial and error methods for the ultimate determination. The mathematical treatment was considerably simplified by Melville (1937) in an attempt to apply the method to the vapour-phase polymerization of methyl methacrylate. The choice of this monomer made it impossible, on account of the complexity of the reaction, to obtain much more than a very rough idea of the life time.

The theory which will be elaborated here is a simplification of a somewhat more complicated theory due to Dickinson (1941). The theory will be applicable to all chain reactions in which the rate obeys the square-root law which has been mentioned. In the case of polymers there is the added difficulty that there is the possibility that the values of the kinetic coefficients will depend on the molecular size of the polymer molecules involved, so that in the simplified theory only an average value for these quantities can be obtained. It is, however, easily practicable to change the molecular size of the particles at will by varying the intensity of the light, since the molecular weight of the product is inversely proportional to the square root of the light intensity. By this method it is obviously possible to investigate any change in the values of the kinetic coefficients with molecular size. By measuring the life time of the active particle at different temperatures it is also possible to separate the energies of activation for the various stages of the reaction, and this will eventually lead to the complete fundamental analysis of the reaction.

One of the main difficulties with this reaction is that the chain length cannot be reduced much below 150. With this length of chain it is highly probable that the kinetic coefficients are relatively insensitive to further changes in molecular size. At much lower chain lengths, say of the order of 10, this is unlikely to be true, and an investigation along those lines might be profitable.

THEORY OF INTERMITTENT ILLUMINATION

In general, it can be stated that if the rate, R, of a photochemical reaction is related to the intensity of absorbed radiation by

$$R \propto \{f(I)\}^n,$$

where $n < 1$, then for intermittent illumination in which the periods of light and dark are very short, i.e. very much less than the life time of the active species, and are equal, the situation is the same as that for using half the intensity, so that

$$R_{\text{short}} \propto \{\tfrac{1}{2}f(I)\}^n.$$

For long periods of light and dark the system is virtually illuminated for only half the time, so that

$$R_{\text{long}} \propto \tfrac{1}{2}\{f(I)\}^n.$$

Hence follows the ratio

$$\frac{R_{\text{short}}}{R_{\text{long}}} = 2^{1-n},$$

and if $n = \frac{1}{2}$, then

$$R_{\text{short}}/R_{\text{long}} = \sqrt{2}.$$

In the polymerization reaction it will be assumed that steps are represented by the following sequence of reactions:

$M \rightarrow P_1$	initiation	$f(I)$
$P_n + M \rightarrow P_{n+1}$	propagation	k_p
$P_n + P_m \rightarrow M_n + M_m$	termination	k_t

where M is a monomer molecule, P_r an active polymer molecule of r units, and M_r the corresponding deactivated polymer. When a steady state has been reached it is obviously true that the stationary state equation will be

$$\frac{d(P)}{dt} = f(I) - k_t(P)_s^2 = 0, \tag{1}$$

where $(P)_s$ is the steady concentration of all the active particles under a steady illumination of $f(I)$ and k_t is the termination velocity coefficient. Hence, from equation (1),

$$(P)_s = \{f(I)/k_t\}^{\frac{1}{2}}. \tag{2}$$

For intermittent illumination the periods of light and dark have to be considered separately. During the light period the variation in (P) will be governed by the equation

$$\frac{d(P)}{dt} = f(I) - k_t(P)^2, \tag{3}$$

and during the dark period by

$$\frac{d(P)}{dt} = -k_t(P)^2. \tag{4}$$

Integration of equation (3) leads to the instantaneous value of the concentration of all the active particles at any time during the light flash, i.e.

$$\left\{\frac{f(I)}{k_t}\right\}^{\frac{1}{2}} \tanh^{-1}\left\{\frac{k_t}{f(I)}\right\}^{\frac{1}{2}}(P) = f(I)\,t + c_1, \tag{5}$$

which, by substitution of the expression given in equation (2), leads to the form

$$\frac{1}{(P)_s}\tanh^{-1}\frac{(P)}{(P)_s} = k_t t + c_2. \tag{6}$$

Integration of equation (4) leads to the value of (P) during the dark period. This will be

$$\frac{1}{(P)} = k_t t + c_3, \tag{7}$$

where $(P)_s$ is the value of (P) under steady illumination, and c_1, c_2 and c_3 are constants of integration.

The average concentration of the active particles during the light period will be given by the expression

$$\begin{aligned}(P)_l &= \frac{1}{t}\int_{\text{over light}} (P)\,dt \\ &= \frac{(P)_s}{t}\int_{T_2}^{T_1} \tanh\{(P)_s(k_t t + c_2)\}\,dt \\ &= \frac{1}{k_t t}\ln\left\{\frac{1-\tanh^2(P)_s(k_t T_2 + c_2)}{1-\tanh^2(P)_s(k_t T_1 + c_2)}\right\}^{\frac{1}{2}} \\ &= \frac{1}{2k_t t}\ln\left\{\frac{1-\{(P)_2/(P)_s\}^2}{1-\{(P)_1/(P)_s\}^2}\right\}. \end{aligned} \tag{8}$$

In the same way the average value of (P) during the dark period is given by

$$(P)_d = \frac{1}{2rk_t t}\ln\left\{\frac{(P)_1}{(P)_2}\right\}^2, \tag{9}$$

where $(P)_1$ and $(P)_2$ denote the values of (P) at the beginning and end of the dark period, t is the duration of the light period and r is the ratio of the dark to the light period.

The average value of (P), $(P)_0$ over all will be defined by

$$(r+1)(P)_0 = (P)_l + r(P)_d. \tag{10}$$

Substituting equations (8) and (9) in equation (10) leads to the expression

$$(r+1)(P)_0 = \frac{1}{2k_t t}\left\{\ln\left[\frac{1-\{(P)_2/(P)_s\}^2}{1-\{(P)_1/(P)_s\}^2}\right]+\ln\left\{\frac{(P)_1}{(P)_2}\right\}^2\right\}. \tag{11}$$

From equation (6) is seen the relationship that

$$\tanh^{-1}\frac{(P)_1}{(P)_s}-\tanh^{-1}\frac{(P)_2}{(P)_s} = k_t t(P)_s = \{f(I)\}^{\frac{1}{2}}\, t k_t^{\frac{1}{2}} = m, \tag{12}$$

and from equation (7)

$$\frac{1}{(P)_2}-\frac{1}{(P)_1} = k_t rt. \tag{13}$$

Equation (12) may be rewritten in logarithmic form to give

$$2\left[m+\ln\frac{1+(P)_2/(P)_s}{1+(P_1)/(P)_s}\right] = \ln\left\{\frac{1-\{(P)_2/(P)_s\}^2}{1-\{(P)_1/(P)_s\}^2}\right\}.$$

Substituting this in equation (11) leads to

$$(r+1)(P)_0 = \frac{1}{k_t t}\left\{m+\ln\left[\frac{1+(P)_2/(P)_s}{1+(P)_1/(P)_s}\right]+\ln\frac{(P)_1}{(P)_2}\right\}. \tag{14}$$

Substituting further the value of $(P)_2$ given by equation (13), then

$$(r+1)(P)_0 = \frac{m}{k_t t}\left\{1+\frac{1}{m}\ln\left[1+\frac{rm}{1+(P)_s/(P)_1}\right]\right\}. \tag{15}$$

Since from equation (12) $m = k_t t(P)_s$, equation (15) will reduce to

$$(r+1)\frac{(P)_0}{(P)_s} = 1+\frac{1}{m}\ln\left[1+\frac{rm}{1+(P)_s/(P)_1}\right]. \tag{16}$$

For the steady state the life time, τ, of the active particle will be defined by

$$\tau = \frac{\text{number of } P \text{ per unit volume}}{\text{number of } P \text{ disappearing per unit volume}}$$

$$= (P)_s/k_t(P)_s^2 = 1/k_t(P)_s,$$

which, from the definition of m, gives

$$\tau = t/m.$$

Hence m is the number of times that the duration of the light flash is greater than the life time of the active polymer.

Figure 7 shows the value of $(P)/(P)_s$ plotted as a function of m, calculated from equations (6) and (7), assuming that the constants of integration can be neglected. It will be seen that for slow rotation of a sector, i.e. when the periods of light and dark are long compared with the life time of the active polymer the value of $(P)/(P)_s$ rises to almost unity during the light period and falls to almost zero during the subsequent dark period. During the succeeding cycles this behaviour is simply repeated. When the periods of light and dark are short in comparison with the life time, i.e. when a fast-rotating sector is used, the value of $(P)/(P)_s$ rises rapidly during the light period and falls less rapidly during the dark period until a state is reached in which the rise during the light period is offset by the fall during the dark period. This results in the rapid establishment of a steady state, which, in figure 7, is when the value of $(P)/(P)_s$ is 0·5, since the ratio of light to dark periods is 1:3 in this case.

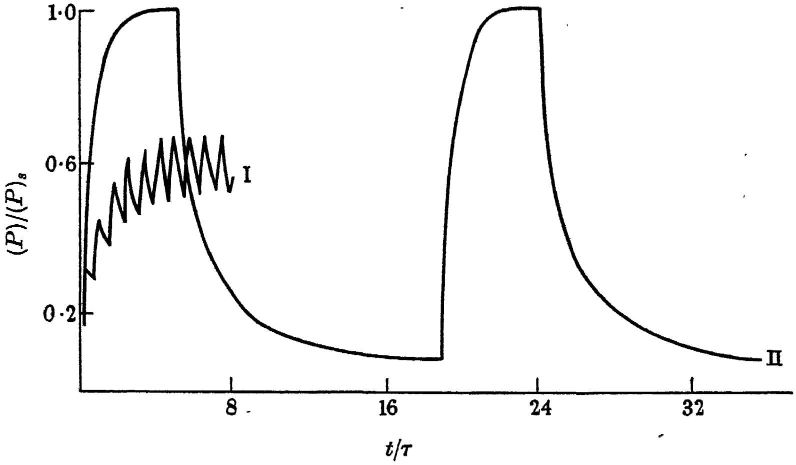


FIGURE 7. Value of $(P)/(P)_s$ as a function of time. Curve I, high-sector speed. Curve II, low-sector speed.

Figure 8 shows the values of $(P)_1/(P)_s$, $(P)_2/(P)_s$ and $(P)_0/(P)_s$ again as a function of m for the ratio of dark to light periods equal to 3. The important part of this figure is the first curve, since it is from this and the others corresponding to it for different values of r that the curves shown in figure 9 are derived. These curves show the variation in the overall concentration of the active species with sector speed for various values of r. For convenience the ordinates have been multiplied by $(r+1)^{\frac{1}{2}}$ so that the curves are all asymptotic to unity. It will be seen that the higher the value of r the more rapid is the descent of the curve, but this advantage is offset by the fact that the overall rate of the reaction is considerably reduced. Indeed, in many cases there is little advantage in using a high ratio of light to dark, and equal periods of light and dark will suffice since the overall change in velocity is of the order of 30 % for the transition from very fast to very slow sector speeds.

As was stated earlier this idea is based on a somewhat similar analysis due to Dickinson (1941) and possesses several advantages over the original theory by Chapman *et al.* (1926). As will be seen from the curves in figure 9, for $r = 1$, i.e. for equal periods of light and dark, the value of $(P)_0/(P)_s$ begins to fall off when $m = 1$, i.e. when the duration of the light flash is equal to the life time of the active particle under steady illumination. Now the rate of the reaction will be proportional to the concentration of the active particles, so that the relative values of the rates under steady and intermittent illumination will follow the same course as the curves for the relative concentrations under the different conditions. This means that the value of the rate will also begin to fall when the value of m is unity, so that the value of the life time of the active species can be stated *definitely* without further calculation. It should be noted that the value of the life time obtained is under steady illumination. Alternatively, the rate of reaction may be plotted against $\log(f(I)^{\frac{1}{2}}t)$, and on the same scale the appropriate curve from figure 9. The separation of these two curves will automatically give the value of $\log(k_t)^{\frac{1}{2}}$, so that the velocity coefficient for the termination reaction may be determined directly.

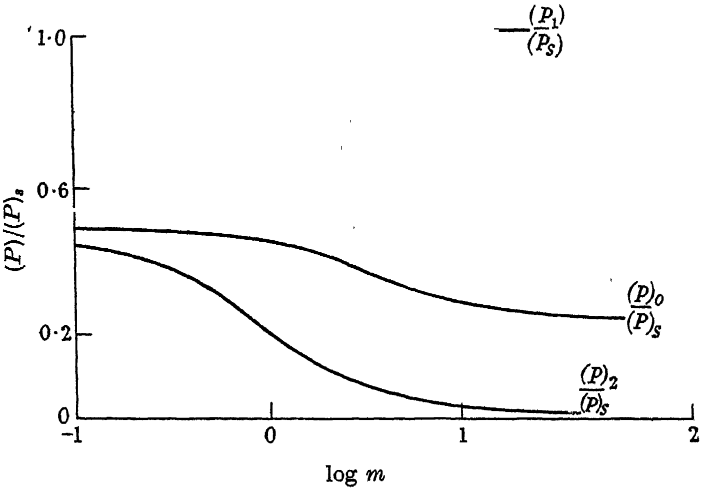


FIGURE 8. Values of $(P)_1/(P)_s$, $(P)_2/(P)_s$, $(P)_0/(P)_s$ as functions of m.

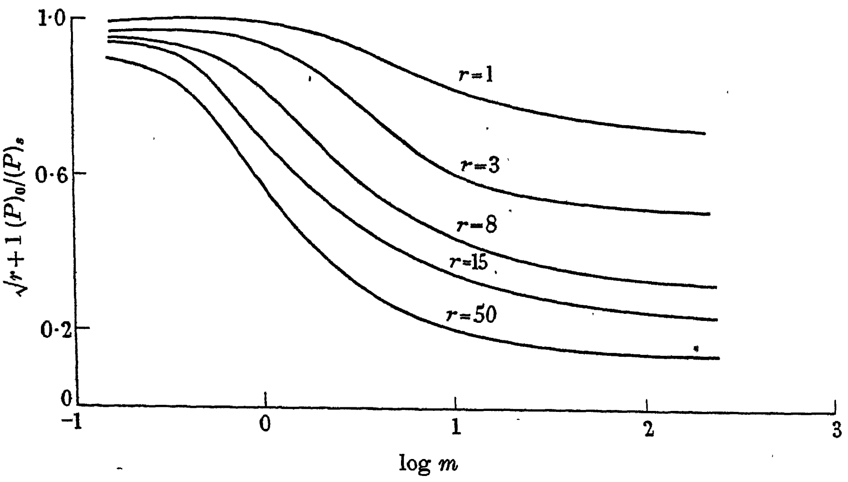


FIGURE 9. Theoretical curves of reaction velocity as functions of m, for various values of r. Curve I, $r=1$; curve II, $r=3$; curve III, $r=8$; curve IV, $r=15$; curve V, $r=50$.

Another method of finding the value of τ eliminates the necessity of drawing the full curves with any degree of accuracy. It is merely necessary to obtain the rate of reaction with the sector rotating at a very high speed and then at some intermediate speed for which the duration of the light flash is known. The ratio of the rates for these two speeds is calculated, and the corresponding value of m found from the theoretical curve which is applicable. Hence the value of τ may be found since m is, by definition, the number of times that the duration of the light flash is greater than the life time. Such a simplification was not used here, since it was of greater interest to see how the theory corresponded to practice, but the method was used later in connexion with the value of the kinetic coefficients in precipitating media, details of which will be published in a later paper. The term $f(I)$ which has been used throughout will, in effect, be the absolute rate of initiation of the chains and, in this case, fortunately, it is possible to deduce this from the benzoquinone inhibition which was described earlier. This naturally limits the applicability of the method to systems in which the evaluation of the primary quantum efficiency can be determined.

Experimental technique

The simplest method of obtaining intermittent illumination is by the use of disks, cut with sectors, which are rotated at a constant speed in the path of the light. This practice has been followed here. The disks were made from galvanized iron sheet, painted black to avoid reflexion effects, and having a diameter of about 14 in. The sectors cut in the disks had a depth of about 2 in. and an opening of either $\frac{1}{4}\pi$ or $\frac{1}{16}\pi$. The former was used when long periods of light and dark were required and the latter for short periods of illumination—down to about 7 msec. This was found to be the most satisfactory method, since it was not found to be possible to control the speed of the motor at excessively low speeds. The sector opening has to be greater than the area of the reaction vessel in order to avoid merely scanning the system.

The sector disk was screwed to a copper bush whose spindle was insulated and connected to the driving axle of a 35 W universal electric motor. The outside of the spindle of the bush was fitted with two pieces of inlaid insulation. Brushes, made from phosphor bronze strip, were constructed to run on the spindle; the system was connected through a 12 V supply to a Veeder Root magnetic counter having a resolution greater than 800/min. The time required for a given number of revolutions was measured on a stop-watch reading to 0·1 sec., and from this the duration of the light flash could be calculated. The speed of the motor could be varied by means of a 'Variac' autotransformer whose output varied from 0 to 250 V. Checks on the speed of the motor were frequently made in order to ensure constant running conditions.

In carrying out a run the following procedure was adopted. Since, as has already been pointed out, the rate of reaction is constant during the initial stages, it is possible to carry out a complete determination of the life time in one run. The sector was allowed to run at a constant speed for 10 to 15 min., and the drop in the meniscus level measured by means of a cathetometer. The speed of the motor was altered and the rate of reaction again determined. This made it possible to cover a considerable number of sector speeds in one run, but obviously the technique can be used only if the reaction is a linear one.

These experiments were carried out at temperatures of 15·9 and 31·4° C to find the values of the energies of activation for the termination and propagation steps of the reaction, since it is assumed that the rate of initiation is independent of temperature. To check the possible effect of molecular size the intensity was reduced by the insertion of a suitable screen, to 0·25 of its original value, which has the effect of doubling the chain length.

Results

The main results of the experiments are set out in table 7.

TABLE 7. RATE OF REACTION COMPARED WITH DURATION OF FLASH

at 15·9° C		at 31·4° C	
time of flash (msec.)	% rate for high sector speeds	time of flash (msec.)	% rate for high sector speeds
10	100	10	100
15·5	100	15	100
20	100	20	100
22	100	35	95·5
35	96·4	90	88·5
45	80·6	250	79·0
180	76·7	350	74·2
240	73·6		

The values are plotted as a function of the time of flash in figure 10. From this curve compared with the theoretical curve and using the method of interpretation suggested, the life time of the active particle is $2{\cdot}25 \times 10^{-2}$ sec. at 15·9° C.

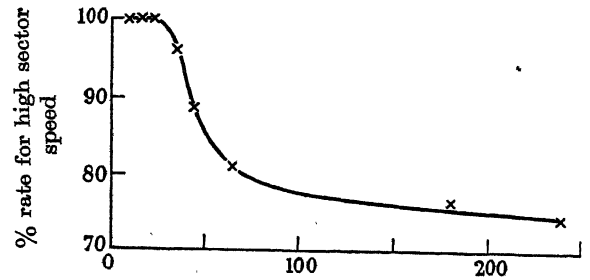


FIGURE 10. Rate of reaction as a function of duration of flash.

As was explained earlier, the equations used in the calculation of the individual constants have to be simplified in comparison with those used by Melville & Gee. It is assumed that for the propagation step the actual value of the velocity coefficient can be written as k_p, but there is no assumption that this coefficient is, indeed, independent of molecular size. Similarly, the velocity coefficient for termination is written as k_t. Thus it will be seen that the results quoted here will be average values for some particular average degree of polymerization. Making these assumptions the steady state equation will be

$$\frac{d(P)}{dt} = f(I) - k_t(P)^2 = 0,$$

where (P) denotes the concentration of all the active particles. The overall rate of reaction will be given by the expression

$$-\frac{d(M)}{dt} = k_p(M)(P) = \frac{k_p(M)\{f(I)\}^{\frac{1}{2}}}{k_t^{\frac{1}{2}}}. \tag{17}$$

Since $f(I)$ is the intensity of the light usefully absorbed it is, in effect, the rate of starting of the chains. It was found that the length of the inhibition period for the addition of 0·0024 g. of benzoquinone was 48 min. Assuming, as previously, that one molecule of the quinone is effective in stopping a chain then the useful light input $f(I)$ is

$$f(I) = 5{\cdot}0 \times 10^{15}\, h\nu/\text{sec.}$$

The final results are required to be in litre, g.mol. and second units, so that this has to be converted to units corresponding to those. The measured capacity of the reaction vessel was 13·0 c.c., and we consider that we are justified in saying that this is the volume of the absorbing medium. Hence

$$f(I) = 6{\cdot}35 \times 10^{-7}\,\text{einstein l.}^{-1}\,\text{sec.}^{-1}.$$

The overall rate of polymerization under this steady illumination was found by direct measurement on the dilatometer and gives that

$$-\frac{d(M)}{dt} = 1{\cdot}1 \times 10^{-4}\,\text{l. g.mol.}^{-1}\,\text{sec.}^{-1},$$

taking the density of vinyl acetate as 0·86.

It is possible to show that under conditions of steady illumination, the following equation is true provided that τ is small (Jones & Melville 1940):

$$\frac{d(P)}{dt} = \frac{(P)}{\tau}. \tag{18}$$

In this case, the value of $d(P)/dt$ is given, which is the rate of starting chains, and, the life time of the active species, so that it is possible to calculate the concentration of the active particles. This gives

$$(P) = 1{\cdot}42 \times 10^{-8}\,\text{g.mol./l.}$$

In order to calculate the velocity coefficient for the growth process equation (17) can be rewritten in the form

$$k_p = \frac{-d(M)/dt}{(P)(M)}.$$

Substituting the relevant values in this, then

$$k_p = 7{\cdot}7 \times 10^{2}\,\text{l.g.mol.}^{-1}\,\text{sec.}^{-1}.$$

Also from equation (18) it is seen that

$$k_t = \frac{k_p^2(M)^2 f(I)}{\{-d(M)/dt\}^2},$$

which, on substitution of the appropriate figures, leads to

$$k_t = 3{\cdot}1 \times 10^9 \,\text{l. g.mol.}^{-1}\,\text{sec.}^{-1}.$$

These experiments were repeated at 31·4° C when it was found that the value of the life of the active polymer had not changed appreciably—the new value of the life time was $2{\cdot}30 \times 10^{-2}$ sec. It was considered that this was within the limits of experimental error to be expected from this type of reaction, so it can be said that the life time does not change. It is obvious from the analysis that has preceded this that if the life time does not change then the value of the velocity coefficient for the termination process also has not changed. This means that the energy of activation for the termination process must be zero, and, therefore, the overall energy of activation which has already been given must be the energy of activation of the propagation step, since it is justifiable to assume the initiation rate to be temperature independent. Since k_p must be defined by

$$k_p = A_p e^{-E/RT}, \tag{19}$$

which gives a value of $1{\cdot}65 \times 10^6$ l. g.mol.$^{-1}$ sec.$^{-1}$ for A_p.

Assuming, as is most likely, that the growth reaction is as a result of a bimolecular collision between the growing polymer and the monomer, the number of such collisions taking place in a normal liquid per sec. is 10^{11} (Moelwyn-Hughes 1940), so that the steric factor will be of the order of 10^{-5}. The value of 10^{11} for the number of collisions per sec. is taken, in this case, since it is believed that the number of collisions between the monomeric molecules and the active end of the chain will be of the same order as that between molecules in a pure liquid. In effect, the motion of the macromolecule as a whole is not considered but only the active part of it, whose motion will not be seriously restricted. In the same way it is obvious that the A factor for the termination reaction will be of the order of 10^9 so that the steric factor will be around 10^{-2}.

From the constants which have been given it is possible to calculate the value of the kinetic chain length. This is done most simply from the relationship

$$\nu_k = \frac{-d(M)/dt}{f(I)},$$

where ν_k is the kinetic chain length. The relevant figures give the value of the chain length as 174. In order to check this conclusively osmometric measurements were made on the polymer in order to obtain the number average molecular weight. This was found to correspond to a chain length of 176.

A similar series of experiments was carried out with a light intensity of $1{\cdot}6 \times 10^{-7}$ einstein l.$^{-1}$ sec.$^{-1}$. In this case the chain length was increased, but the life time did not increase in quite the same ratio. Table 8 gives the results of both series for comparative purposes.

TABLE 8. THE KINETIC COEFFICIENTS AT 15·9° C

$f(I)$ (einstein l.$^{-1}$ sec.$^{-1}$)	$6{\cdot}35 \times 10^{-7}$	$1{\cdot}60 \times 10^{-7}$
τ (sec.)	$2{\cdot}25 \times 10^{-2}$	$5{\cdot}10 \times 10^{-2}$
$-d(M)/dt$ (g.mol. l.$^{-1}$ sec.$^{-1}$)	$1{\cdot}10 \times 10^{-4}$	$5{\cdot}51 \times 10^{-5}$
k_p (l.mol.$^{-1}$ sec.$^{-1}$)	$7{\cdot}7 \times 10^{2}$	$6{\cdot}7 \times 10^{2}$
k_t (l.mol.$^{-1}$ sec.$^{-1}$)	$3{\cdot}1 \times 10^{9}$	$2{\cdot}5 \times 10^{9}$
k_t/k_p	$4{\cdot}0 \times 10^{6}$	$3{\cdot}8 \times 10^{6}$
ν_k	174	339
ν_m	176	352
steric factor (propagation)	10^{-5}	10^{-5}
steric factor (termination)	10^{-2}	10^{-2}

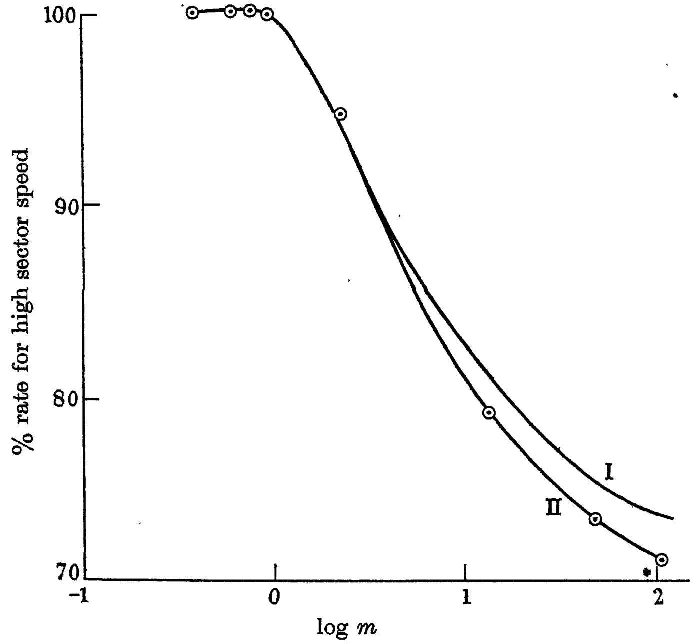


FIGURE 11. Experimental and theoretical curves. Curve I, theoretical. Curve II, experimental.

There appears to be some reason to expect a slight variation in the values of the kinetic coefficients, as the degree of polymerization varies. This is compatible with the idea that the macromolecules are extended in solution. If the original molecular model of a spherical molecule with an active patch had been correct, it is easy to see that the order of the change would have been considerably greater. Thus an extended molecule is seen whose active end is subjected to only slightly restricted motion.

Figure 11 shows a normalized experimental curve (curve II) with the theoretical curve (curve I). It will be seen that except in the lower regions there is very good agreement. It may be that the discrepancy is due to the fact that the theory of polymer reactions has had to be very much simplified in order to allow of the formulation of the theory of intermittent illumination. In the lower region, i.e. with slower sector speeds, if the degree of polymerization increased it is possible that there is a drop in the value of the termination coefficient, this would account for the discrepancy.

The most interesting of the results appears to be the comparatively high value of the steric factor for termination as compared with that for the propagation step. Evans (1946) has attempted to explain this on the basis of the transition state theory. For this he assumes that the monomeric unit in the free state is capable of oscillating more freely than one incorporated in the polymer molecule. From this premise he has been able to correlate these results with the theory and derive results which are of a reasonable order of magnitude considering the nature of the reaction.

The authors have to thank Mr C. R. Masson, B.Sc., for the osmometric results quoted, and one of us (G.M.B.) is indebted to Distillers' Co. Ltd. for financial assistance (1945–6).

REFERENCES

Bolland, J. L. & Melville, H. W. 1938 *Proc. Rubber Tech. Conf.* Preprint no. 90.
Breitenbach, J. W. & Raff, R. 1939 *Ber. dtsch. chem. Ges.* **69**, 1107.
Chapman, D. L., Briers, F. & Walters, E. 1926 *J. Chem. Soc.* p. 562.
Cruickshank, J., Masson, C. R., Melville, H. W. & Menzies R. F. 1946 *Nature,* **157**, 74.
Cuthbertson, A. C., Gee, G. & Rideal, F. K. 1937 *Nature,* **140**, 889.
Dickinson, R. G. 1941 *Photochemistry of gases,* p. 200. Leighton and Noyes (Reinhold).
Evans, M. G. 1946 (In the Press.)
Foord, S. G. 1940 *J. Chem. Soc.* p. 48.
Fuoss, R. M. & Mead, E. J. 1943 *J. Phys. Chem.* **47**, 59.
Goldfinger, G., Skeist, I. & Mark, H. 1943 *J. Phys. Chem.* **47**, 578.
Jones, T. T. & Melville, H. W. 1940 *Proc. Roy. Soc.* A, **175**, 392.
Marchlewski, L. & Moroz, A. 1924 *Bull. Soc. chim. Paris,* **35**, 473.
Melville, H. W. 1937 *Proc. Roy. Soc.* A, **163**, 511.
Melville, H. W. & Gee, G. 1944 *Trans. Faraday Soc.* **40**, 240.
Moelwyn-Hughes, E. A. 1940 *Physical chemistry,* p. 72. Camb. Univ. Press.
Norrish, R. G. W. & Brookman, E. F. 1937 *Proc. Roy. Soc.* A, **163**, 205.
Schulz, G. V. 1941 *Z. Elektrochem.* **47**, 265.
Starkweather, H. W. & Taylor, G. B. 1930 *J. Amer. Chem. Soc.* **52**, 4708.
Vernon, A. A. & Taylor, H. S. 1931 *J. Amer. Chem. Soc.* **53**, 2527.

Determination of the velocity coefficients for polymerization processes

II. The photosensitized polymerization of vinyl acetate by benzoyl peroxide

By G. M. BURNETT AND H. W. MELVILLE, F.R.S.

Chemistry Department, University of Aberdeen

(*Received* 7 *November* 1946—*Read* 6 *February* 1947)

The photopolymerization of vinyl acetate has been found to be very much accelerated by the presence of small quantities of benzoyl peroxide. This is considered to be the result of a photochemical decomposition of the peroxide, to give free radicals such as phenyl and benzoate. The dependence of the rate of polymerization on the peroxide concentration suggests that there is a considerable inhibition of the reaction due to the presence of the peroxide. The value of the intensity exponent approaches unity as the concentration of the peroxide is increased, and the formal analysis of the reaction suggests that the inhibition is due to the peroxide molecules themselves. The analysis of the induction period and the dark period shows that these are identical and therefore real, so that it can be concluded that they depend essentially on the nature of the initiation step. The value of the initiation velocity coefficient is calculated, and the energy of activation for this step is found to be 8·4 kcal./g.mol.

INTRODUCTION

The results of the direct photopolymerization of vinyl acetate suggest, although no concrete proof is available, that the chain carrier is a free radical. By adding benzoyl peroxide to the monomer and illuminating at temperatures at which the thermal catalyzed reaction is negligible, it was found that a photosensitized reaction took place. If a kinetic similarity between a radical reaction and the photo-excited reaction can be established, then the proof of the radical nature of the chain carrier in the latter case is without doubt. It is evident, on account of this sensitized reaction, that such a method is not straightforward. The thermal breakdown of benzoyl peroxide gives rise to free radicals probably by the scheme (Hey & Waters 1935)

$$C_6H_5CO.O.O.OC.C_6H_5 \rightarrow C_6H_5— + C_6H_5.COO— + CO_2.$$

One or other of these free radicals is then capable of initiating a polymerization chain. If a similarity in kinetics can be established between the thermal catalyzed and the sensitized reaction, it is evident that the latter is also a free radical reaction. Establishment of a like similarity between the direct photochemical and the sensitized reaction will again show an identity of mechanism for these processes. It is not possible to make such a comparison directly, since the catalyzed reaction has a pronounced induction period.

The only work on the photodecomposition of benzoyl peroxide suggests that this reaction may have as a first step the formation of similar radicals to the thermal case, but since the investigation was carried out in the solid phase no conclusive evidence of this is available (Fichter & Schnider 1930). Diacetyl peroxide has also been investigated, and the results obtained in solution and with the pure liquid suggest the formation of methyl and acetyl radicals, as the main products of the decomposition are ethane and methane, approximately in the ratio 1:1, in the case of the pure liquid (Walker & Wild 1937). By analogy a similar reaction might be possible for benzoyl peroxide.

Experimental

Benzoyl peroxide, which had been purified in the usual way, was made up as a standard solution in benzene. A known volume of this solution was introduced into a silica reaction tube and the solvent was removed in the vacuum from a water pump. The tubes were then sealed to a length of silica tube provided with a ground joint and attached to a suitable manifold. Vinyl acetate monomer, purified by fractional distillation and thoroughly freed from air, was distilled into the tubes under high vacuum and the tubes sealed off.

The rate of polymerization was followed dilatometrically, using a cathetometer reading to 0·01 mm.

The photosensitized reaction

With low concentrations of peroxide and temperatures of the order of 30° C the rate of the thermal reaction is negligible. The rates of polymerization using peroxide concentrations from 6×10^{-4} to 6×10^{-1} g.mol./l. were determined. At higher concentrations the thermal catalyzed reaction ceased to be negligible, so that the investigation of the sensitized reaction beyond this point was not made. Figure 1 shows a typical run in which it will be noted that there is a clearly defined induction period, after which the reaction is apparently of zero order with respect to the monomer concentration. When the source of illumination is removed the reaction continues for a short period. The fact that the rate curve for the dark period is the inverse of that of the induction period makes it certain that both of these observations are real. Had the reaction been non-isothermal, then during the induction period the contraction due to the polymerization would have been offset to some extent by the expansion of the liquid itself. This would have given the appearance of an induction period. The reverse would be true during the dark period, so that the contraction during this period would be due to the polymerization and the contraction of the solution cooling down. The fact that the two stages are superposable ensures that there is no non-isothermal effect. For comparison purposes the direct photopolymerization, which was measured under the same conditions, is also plotted in figure 1 curve II and is obviously negligible.

At very low peroxide concentrations the rate of polymerization is proportional, for a constant light intensity, to the square root of the concentration. This is clearly seen from the figures in table 1.

At higher value of the peroxide concentration this simple dependence disappears, and the rate of polymerization becomes less dependent on the concentration of peroxide. After a certain concentration is reached the rate of polymerization begins to fall slightly as the peroxide concentration is increased. Table 2 shows the results obtained with the higher catalyst concentrations, in which this behaviour is demonstrated. The falling off in the rate at this stage must mean that the peroxide molecules or one or other of the radicals formed must be capable of stopping the chains. The removal of the radicals or molecules in this way would lower the velocity of polymerization, first, by the premature removal of potential chain initiators and, secondly, by the shortening of the reaction chain. There is also the possibility of the removal of possible initiators by the formation of diphenyl or phenyl benzoate.

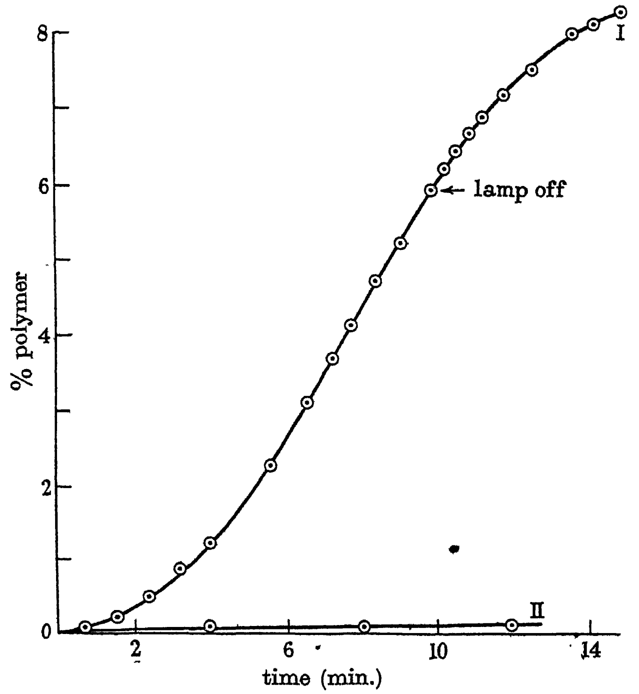


FIGURE 1. The photosensitized polymerization of vinyl acetate, with benzoyl peroxide, at 30° C.

TABLE 1. EFFECT OF LOW PEROXIDE CONCENTRATIONS ON THE RATE OF REACTION

peroxide concentration C (g.mol./l.)	max. rate R (% per sec.)	R/C	$R/C^{\frac{1}{2}} \times 10^{-2}$
$6{\cdot}0 \times 10^{-4}$	0·00113	1·87	4·64
$1{\cdot}2 \times 10^{-3}$	0·00160	1·33	4·66
$2{\cdot}0 \times 10^{-3}$	0·00206	1·03	4·62
$4{\cdot}8 \times 10^{-3}$	0·00350	0·73	5·04

A somewhat similar situation was observed by Schulz (1941) in the case of the polymerization of both styrene and methyl methacrylate using free radicals formed by the thermal decomposition of tetraphenylsuccinodinitrile. In this case there is also a type of inhibition at high radical concentrations which he attributes to the addition of the dinitrile molecule to the double bond of the monomer. This will again result in the removal of radicals which have the ability to initiate chains.

The intensity of radiation was varied by means of suitable perforated screens, of known transmission. The rates of reaction under the full intensity and under the reduced intensity were measured separately in the usual way. The dependence of the rate of reaction on the intensity of illumination is a sensitive test of the occurrence of inhibition, so that the intensity exponent was measured over a considerable range of peroxide concentration, in order that more information on the nature of the inhibition could be obtained. Table 3 shows the values of the intensity exponent obtained in these experiments.

TABLE 2. EFFECT OF HIGH PEROXIDE CONCENTRATIONS ON THE RATE OF REACTION

peroxide concentration (g.mol./l.) $\times 10^2$	rate (% per sec.) $\times 10^2$
9·5	1·59
19·0	2·15
29·0	2·52
38·0	2·46
46·0	2·31
58·0	2·23

TABLE 3. THE INTENSITY EXPONENT

Screen transmission = 0·18.

peroxide concentration (g.mol./l.) ($\times 10^2$)	log (R_r/R_f)	intensity exponent
4·76	−0·337	0·48
9·52	−0·421	0·55
19·0	−0·440	0·59
28·6	−0·462	0·62
38·1	−0·650	0·88
57·6	−0·663	0·89

R_f is the rate of polymerization under full intensity, and R_r under the reduced intensity.

From these results it is evident that at the higher peroxide concentrations there is a side reaction which has a marked effect on the value of the intensity exponent. The trend of this effect may be somewhat surprising, but this point will be discussed in detail later. Under the conditions of reduced intensity the length of the induction period remains constant within the limits of experimental error. This is to be expected, since it depends solely on the value of the velocity coefficient for the initiation reaction, which would be unaffected by the change in intensity.

It is not possible to vary the temperature over a very wide range, since at the higher temperatures the thermal catalyzed reaction will begin to have a considerable effect. In order to achieve as high temperatures as possible the concentration of the peroxide was maintained throughout at $4{\cdot}8 \times 10^{-2}$ g.mol./l. This also reduces the inhibition reaction to negligible proportions. The results of these experiments have been plotted in the usual way in figure 2, from which the overall energy of activation is found to be 8·7 kcal./g.mol. It is not unjustifiable to assume that the energy of activation for the propagation process is unchanged from the bulk polymerization, i.e. 4·40 kcal./g.mol., and the energy of activation for the termination step is zero, then the energy of activation for the initiation process can be found as follows. The overall activation energy, E_0, for a reaction of this type, i.e. with mutual termination, is given by

$$E_0 = \tfrac{1}{2}E_i + E_p,$$

where E_i is the energy of activation for the initiation process and E_p for propagation. Hence, substituting the appropriate values gives 8·6 kcal./g.mol. for E_i.

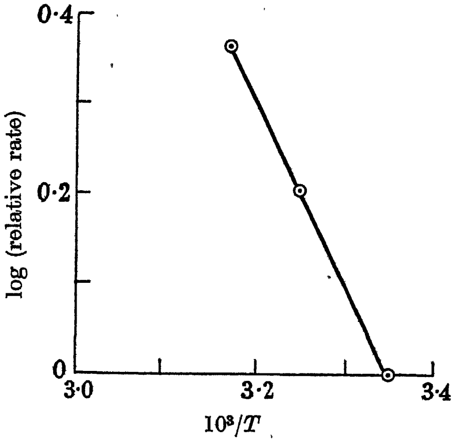


Figure 2. Overall energy of activation for sensitized reaction.

The velocity coefficient for initiation

Since the reaction shows a well-defined induction period it is possible to compute the value of the velocity coefficient for the initiation reaction directly. For the initial period the following equation must govern the concentration of the free radicals:

$$\frac{d(R)}{dt} = f(I) - k_i(M)(R), \tag{1}$$

where (R) is the radical concentration, k_i the velocity coefficient for the initiation reaction, (M) the concentration of the monomer, and $f(I)$ a function of the light intensity. Integration of this leads to

$$-\frac{1}{k_i(M)} \ln\{f(I) - k_i(H)(R)\} = t + c, \tag{2}$$

and the introduction of the condition that $(R) = 0$, when $t = 0$, gives the value of the constant of integration, c,

$$c = -\frac{1}{k_i(M)}\ln\{f(I)\}.$$

Substitution of this in equation (2) and simplifying gives

$$\frac{f(I)}{f(I) - k_i(M)(R)} = e^{-k_i(M)t}, \tag{3}$$

which can be written in the form

$$k_i(M)(R) = f(I)\{1 - e^{-k_i(M)t}\}. \tag{4}$$

The rate of reaction, r, will be proportional to the concentration of the radicals, (R), so that

$$r = \frac{Kf(I)}{k_i(M)}\{1 - e^{-k_i(M)t}\}.$$

The maximum rate of reaction, r_0, will be defined by setting t very large so that the exponential term disappears, and

$$r_0 = \frac{Kf(I)}{k_i(M)}.$$

Hence

$$\frac{r}{r_0} = 1 - e^{-k_i(M)t},$$

or

$$\ln\frac{r_0 - r}{r_0} = -k_i(M)\,t.$$

Hence if $\log(r_0 - r)/r_0$ is plotted against time the result ought to be a straight line whose slope will give the value of $k_i(M)$ and hence the value of k_i. Table 4 gives the rate of reaction with the time, and from the figures it is evident that the linear relationship will hold. The half-life of the induction period is 201 sec., so that the value of $k_i(M)$ will be $4{\cdot}97 \times 10^{-3}$ sec.$^{-1}$, from which, since (M) is 10 g.mol./l.,

$$k_i = 4{\cdot}97 \times 10^{-4}\,\text{l.g.mol.}^{-1}\,\text{sec.}^{-1}.$$

TABLE 4. RATE OF REACTION OVER THE INDUCTION PERIOD

time (sec.)	rate, r (% per sec.)	r/r_0	$\log(1 - r/r_0) \times 10$
40	0·0020	0·141	0·93
90	0·0039	0·278	0·85
140	0·0052	0·366	0·79
190	0·0065	0·457	0·73
240	0·0078	0·715	0·64
340	0·0100	0·715	0·44

$r_0 = 0{\cdot}0140$ % per sec.

TABLE 5. DEPENDENCE OF k_i ON TEMPERATURE

temperature (°C)	26·4	33·3	41·7
k_i (l.g.mol.$^{-1}$ sec.$^{-1} \times 10^4$)	3·64	4·97	7·70

Using a similar method the value of k_i was found at different temperatures and the values are given in table 5. Using the usual logarithmic plot this leads directly to the value of the energy of activation for the initiation process, which is 8·6 kcal./g.mol., which is in good agreement with the result obtained by the indirect method. Since k_i may be defined by

$$k_i = A_i e^{-E/RT},$$

it is possible to calculate the A factor making use of the results which have been obtained. This gives

$$A_i = 5{\cdot}9 \times 10^2 \,\text{l.g.mol.}^{-1}\,\text{sec.}^{-1},$$

and hence it is possible to compute the steric factor. If the liquid is taken as being normal then the number of binary collisions per sec. will be of the order of 10^{11}, so that the steric factor is in the neighbourhood of 10^{-8}.

THE DARK REACTION

After the source of radiation is removed the reaction carries on for some time. As has been stressed this is not due to any non-isothermal effect. The cause of this effect appears to be the inherent inactivity of the radicals produced towards the monomeric molecules. During the dark period the concentration of the radicals will be given by

$$\frac{d(R)}{dt} = -k_i(M)(R).$$

An analysis similar to that shown in the last section will give that the rate, r, at any time, t, is related to the rate, r_0, at zero time by the equation

$$\ln\frac{r_0}{r} = k_i(M)\,t.$$

Table 6 shows the results obtained, and again it will be seen that the plot of log (r_0/r) against time is linear. The value of the half-life of the reaction is 200 sec., so that

$$k_i = 5 \times 10^{-4}\,\text{l.g.mol.}^{-1}\,\text{sec.}^{-1}.$$

This is in good agreement with the previous result, from the half-life of the induction period.

TABLE 6. THE DARK REACTION

time (sec.)	rate, r (% per sec.)	r/r_0	log $(r/r_0) \times 10$
20	0·0129	0·92	0·96
40	0·0122	0·87	0·94
60	0·0112	0·80	0·90
80	0·0106	0·76	0·88
120	0·0095	0·68	0·83
160	0·0078	0·56	0·75
220	0·0067	0·48	0·68
260	0·0057	0·41	0·61
300	0·0049	0·35	0·54
360	0·0038	0·27	0·43

THE MOLECULAR WEIGHT OF THE POLYMER

Since there is no real interest in the absolute value of the molecular weight the relative degrees of polymerization were measured viscosimetrically to ascertain whether there is appreciable inhibition. Table 7 gives the results. The increase in the concentration of the catalyst would lower the molecular weight of the product in the normal course of events, since the chain length is inversely proportional to the square root of the concentration. Hence had the reaction been normal the third column of table 7 ought to be constant, but this is evidently not so. Since the value of the products decrease throughout, the value of the chain length must be less than the rate of starting would indicate, so that it can be concluded that there is an appreciable inhibition effect. Similar results were obtained by Schulz for the free radical initiated polymerization of styrene.

TABLE 7. MOLECULAR WEIGHT DETERMINATION

peroxide concentration C	mol. wt. M	$M \times C^{\frac{1}{2}}$	$M \times C$
$5{\cdot}0 \times 10^{-4}$	76,500	1780	38·25
$1{\cdot}5 \times 10^{-3}$	37,000	1425	55·50
$6{\cdot}0 \times 10^{-3}$	11,000	855	66·00
$2{\cdot}0 \times 10^{-2}$	5,200	728	104·0

FORMAL ANALYSIS OF THE REACTION

For the sake of completeness two possible mechanisms will be analyzed; first, considering the radical and, secondly, the peroxide molecule as being responsible for the inhibition.

I. Assuming that the radical inhibits the reaction:
The reaction sequence for this mechanism would be

$$\begin{array}{ll} B+h\nu\rightarrow 2R & f(I) \\ M+R\rightarrow P_1 & k_i \\ P_1+M\rightarrow P_2 & k_{p_1} \\ \cdots\cdots\cdots\cdots & \\ P_n+M\rightarrow P_{n+1} & k_{p_n} \\ P_n+P_s\rightarrow M_n+M_s & k_{m_n}, k_{m_s}. \\ P_n+R\rightarrow M_n & k_{f_n} \end{array}$$

Only the reaction after the induction period is completed will be considered, since it is then possible to apply stationary state equations for the solution of the problem.

(1) When the inhibition by the radicals is negligible the terms involving k_f may be omitted and it is then easy to show that

(*a*) rate $(\text{peroxide conc.})^{\frac{1}{2}}$,
(*b*) rate $(f(I))^{\frac{1}{2}}$.

(2) When the inhibition by the radicals is comparable with the termination by the mutual reaction, then the stationary state equations take the form

$$\frac{d(P_1)}{dt} = k_i(M)(R) - k_{p_1}(P_1)(M) - k_{t_1}(P_1)(R) - k_{m_1}(P_1)\Sigma k_{m_s}(P_s) = 0,$$

and, generally,

$$\frac{d(P_n)}{dt} = k_{p_{n-1}}(P_{n-1})(M) - k_{p_n}(P_n)(M) - k_{t_n}(P_n)(R) - k_{m_n}(P_n)\Sigma k_{m_s}(P_s) = 0.$$

Summing these equations to infinity gives the result

$$k_i(M)(R) - (R)\Sigma k_{t_s}(P_s) - \{\Sigma k_{m_s}(P_s)\}^2 = 0.$$

Now, as before, write $\delta k_p = k_{m_s}$, and further postulate the relationship $k_{t_s} = \beta k_{p_s}$, so that

$$\delta^2\{k_{p_s}(P_s)\}^2 + \beta(R)\Sigma k_{p_s}(P_s) - k_i(M)(R) = 0,$$

from which $$\Sigma k_{p_s}(P_s) = \frac{-\beta(R) + \{\beta^2(R)^2 + 4k_i(M)(R)\delta^2\}^{\frac{1}{2}}}{2\delta^2}.$$

If, as is likely, $\beta^2(R)^2 \leqslant 4k_i(M)(R)\delta^2$, this will reduce to

$$\Sigma k_{p_s}(P_s) = \left[\frac{\{k_i(M)(R)\}^{\frac{1}{2}}}{\delta} - \frac{\beta(R)}{2\delta^2}\right].$$

For very long chains the rate of polymerization will be given by

$$\text{rate} = (M)\left[\frac{\{k_i(M)(R)\}^{\frac{1}{2}}}{\delta} - \frac{\beta(R)}{2\delta^2}\right].$$

Since the value of (R) must be proportional to some power of (B), the concentration of the peroxide, and also to $f(I)$, it is obviously true that the rate is no longer dependent on the square root of either of these two quantities.

(3) To investigate the extreme case in which the termination by the radicals is very great, one can eliminate from the steady state equations the terms involving k_m to give

$$\frac{d(P_1)}{dt} = k_i(M)(R) - k_{p_1}(P_1)(M) - k_{t_1}(P_1)(R) = 0,$$

$$\frac{d(P_2)}{dt} = k_{p_1}(P_1)(M) - k_{p_2}(P_2)(M) - k_{t_2}(P_2)(R) = 0,$$

and, in general,

$$\frac{d(P_n)}{dt} = k_{p_{n-1}}(P_{n-1})(M) - k_{p_n}(P_n)(M) - k_{t_n}(P_n)(R) = 0,$$

from which it is easy to show that

$$k_{p_n}(P_n) = \frac{k_i(M)^n (R) \prod\limits_{s=1}^{n} k_{p_s}}{\prod\limits_{s=1}^{n} \{k_{p_s}(M) + k_{t_s}(R)\}}$$

$$= \frac{k_i(R)}{\prod\limits_{s=1}^{n} \left\{1 + \frac{k_{t_s}(R)}{k_{p_s}(M)}\right\}}$$

$$= \frac{k_i(R)}{\{1 + \beta(R)/(M)\}^n}.$$

Hence the rate of reaction will be

$$\text{rate} = k_i(M)(R) \left\{\frac{1}{1+\beta(R)/(M)} + \frac{1}{\{1+\beta(R)/(M)\}^2} + \dots\right\}.$$

Since the part in the bracket is a geometrical progression it can be summed to give

$$\text{rate} = k_i(M)(R)\left\{\frac{1/\{1+\beta(R)/M\}}{1-1/\{1+\beta(R)/(M)\}}\right\}$$

$$= \frac{k_i(M)^2}{\beta},$$

which is obviously independent of both the concentration and the value of (I), i.e. the intensity exponent would tend to zero.

II. Assuming that the peroxide molecule acts as an inhibitor:

(i) As before for low values of the peroxide concentration it is easy to show that the rate of polymerization is proportional to the square root of the concentration and of the light intensity.

(ii) When the inhibition by the peroxide is of the same order as the chain breaking by the termination process a similar set of equations to those of section (I(2)) above will be seen except that (R) will be replaced by (B), the benzoyl peroxide concentration. The solution of these equations in the usual way leads to

$$\text{rate} = (M)\left\{\frac{\{k_i(M)(R)\}^{\frac{1}{2}}}{\delta} - \frac{\beta(B)}{2\delta^2}\right\},$$

which is obviously the same equation as was obtained previously except that the intensity dependence will remain virtually unchanged.

(iii) When the inhibition by the peroxide is very great, there is a marked difference from the result given in section (I(3)), and the analysis will therefore be given in detail.

The stationary state equations will have the form

$$\frac{d(P_1)}{dt} = k_i(M)(R) - k_{p_1}(P_1)(M) - k_{t_1}(P_1)(B) = 0,$$

$$\frac{d(P_n)}{dt} = k_{p_{n-1}}(P_{n-1})(M) - k_{p_n}(P_n)(M) - k_{t_n}(P_n)(B) = 0,$$

so that, as before,
$$k_{p_n}(P_n) = \frac{k_i(R)}{\{1+\beta(B)/(M)\}^n},$$

and hence the rate will be given by the expression

$$\text{rate} = \frac{k_i(M)^2\,(R)}{\beta(B)}.$$

From this it must be inferred that the rate of reaction becomes proportional to the first power of the light intensity and also inversely proportional to some power of the peroxide concentration which has a maximum of $\frac{1}{2}$. It would appear, therefore, that the second case fits the facts which have been presented so that one may conclude that in this reaction the first step is the decomposition of the peroxide into free radicals, which are capable of initiating chains of polymerization and that the peroxide molecules at high concentrations are responsible for the inhibition of the reaction.

From the similarity of the kinetics of the reaction of both the thermal catalyzed and the photochemical reactions it can be taken as forming a link between the two reactions. From the recent work on the free radical polymerization of styrene, using free radicals such as diphenyl methyl nitrile, $(\mathrm{C_6H_5})_2\,.\,\overset{|}{\mathrm{C}}\,.\,\mathrm{CN}$, it has been shown that the rates of polymerization and molecular weights obtained are of the same order as with benzoyl peroxide catalysis, so that there is substantial evidence that the latter is free radical in character. The sensitized reaction presents features which are similar to the thermal reaction in that there is a well-defined induction period, and is, at the same time, similar to the photoreaction in that the order of the main reaction is zero. Hence it may be deduced that the fundamental characteristics of the three reactions are the same, and that, in each case, the chain carrier is a free radical.

One matter that has still to be decided is whether the initial photochemical activation of the vinyl acetate gives rise to the production of a diradical or whether the molecule is split into two radicals. In paper I it is shown that *one* molecule of polymer is produced for each molecule of vinyl acetate activated. This result can most readily be explained by supposing that a diradical is in fact produced and that termination at each end involves a disproportionization reaction. This result would, however, also be obtained if the vinyl acetate were dissociated and each radical grew separately, the polymer radicals finally disappearing by combination. There is no evidence to suppose that fission of vinyl acetate does in fact occur. For example in the gas-phase photopolymerization of vinyl acetate (Tuckett & Melville 1947) it can be shown that the molecule is simply activated, for the energy of excitation may readily be removed by collision with argon, helium or hydrogen. This would appear to indicate that absorption of a quantum does not lead to decomposition into free radicals.

A more certain method of deciding the issue is as follows: Figure 3 shows the effect of benzoquinone on the course of the polymerization catalyzed by thermally decomposing benzoyl peroxide—curves II and IV. Curve I shows the course of the direct photo-reaction of such a velocity as to be equal to that of the catalyzed reaction at the end of the induction period. Curve III shows the calculated effect of adding precisely the same weight of benzoquinone. It will be seen that whereas the direct photo-reaction is completely inhibited by benzoquinone the benzoyl peroxide reaction is simply retarded although the quinone is removed and the reaction velocity approaches that in absence of added quinone. It is certain that the peroxide catalyzed reaction involves monoradicals only. It would appear therefore that the direct reaction must involve the production of diradicals, since it has already been shown that both reactions are of the free radical type. Although it is comparatively easy to formulate a structure representing this interaction of a diradical with benzoquinone it is not clear precisely how a monoradical will interact with the quinone. It may be that quinone adds on to the end of the monoradical and the radical produced is inactive towards vinyl acetate, the radical finally disappearing by reaction with another of its kind. Alternatively the quinone might be incorporated in this chain thus forming an interpolymer with the vinyl acetate.

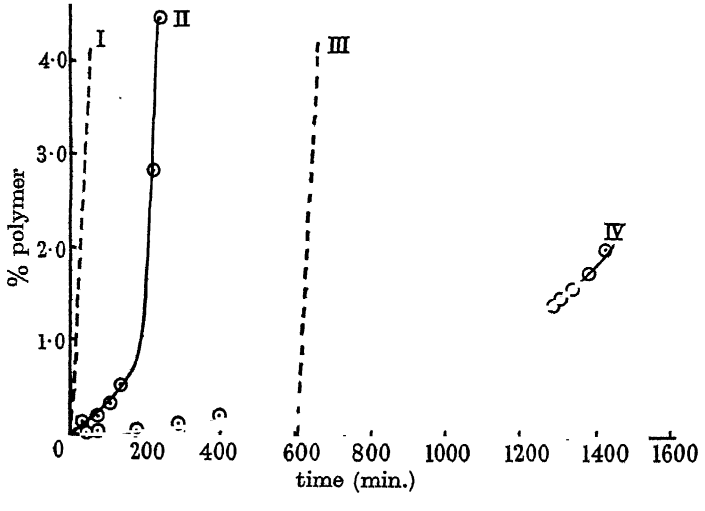
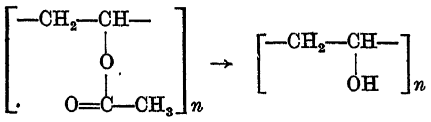


FIGURE 3. Effect of benzoquinone on (*a*) benzoyl peroxide catalyzed reaction at 51·7° C (curve IV), (*b*) direct photopolymerization (curve III).

This disappearance in the action of benzoquinone in inhibiting or retarding polymerization is not confined to vinyl acetate. Similar experiments with styrene and with methyl methacrylate (M. H. Mackay & W. F. Watson unpublished) show the same discrimination. That is, in these reactions involving a monoradical there is retardation, in these reactions—photochemical and thermal—believed to involve diradicals there is inhibition. Consequently it would seem that the use of benzoquinone may provide a useful method for deciding whether monoradicals or diradicals are involved in polymerization reactions.

THE GELVA CYCLE

By analogy with ordinary esters it would be expected that the hydrolysis of polyvinyl acetate would result in the formation of the corresponding alcohol, i.e.



$$\left[\begin{array}{c}-CH_2-\underset{\substack{|\\ O\\ |\\ O=C-CH_3}}{CH}-\end{array}\right]_n \rightarrow \left[\begin{array}{c}-CH_2-\underset{\substack{|\\ OH}}{CH}-\end{array}\right]_n$$

polyvinyl acetate polyvinyl alcohol

The reacetylation of the alcohol should then give the original ester. An investigation of the cycle showed that, in the case of catalytically produced polyvinyl acetates, hydrolysis brought about a decrease in the chain length, so that the reacetylated polymer had a degree of polymerization about half that of the original (McDowell & Kenyon 1940). Subsequent hydrolysis and reacetylation cycles had no further effect on the chain length. It appears, therefore, that the chain rupture takes place at only one point, thus suggesting that there is in the chain a less stable link than the normal carbon-carbon link. McDowell & Kenyon suggest that oxygen from the peroxidic catalyst used or acetaldehyde from the monomer is introduced into the polymer chain at a limited number of points. In photochemically produced polymer formed *in vacuo* the first of these possibilities is avoided, so that a Gelva cycle was performed on a sample of polymer produced in this way.

The hydrolysis and acetylation procedure was the same as that used by the above authors. After the regenerated polymer had been thoroughly dried the molecular weight was found viscosimetrically. The results are shown in table 8. These results show quite conclusively that there is no change in the degree of polymerization in the case of the photochemically produced polymer. This would tend to show that the catalytically produced polymer has a hydrolysable link due to the catalyst used in the production of the polymer.

TABLE 8. GELVA CYCLE ON PHOTOPOLYMER

	intrinsic viscosity	chain length	mol. wt.
original polymer	0·92	354	30,500
after cycle	0·94	362	31,000

One of us (G.M.B.) is indebted to Distillers' Co. Ltd. for financial assistance (1945–6).

REFERENCES

Fichter, F. & Schnider, A. 1930 *Helv. Chim. Acta*, **13**, 1428.
Hey, Q. H. & Waters, W. A. 1937 *Chem. Rev.* **21**, 169.
McDowell, W. H. & Kenyon, W. O. 1940 *J. Amer. Chem. Soc.* **62**, 415.
Schulz, G. V. 1941 *Z. Elektrochem.* **47**, 265.
Tuckett, E. & Melville, H. W. 1947 *J. Chem. Soc.* (in the Press).
Walker, O. J. & Wild, G. L. E. 1937 *J. Chem. Soc.* p. 1132.

Determination of the velocity coefficients for polymerization processes

III. Effect of environment on polymerization processes

By G. M. Burnett and H. W. Melville, F.R.S.

Chemistry Department, University of Aberdeen

(*Received* 7 *November* 1946—*Read* 6 *February* 1947)

The kinetics of the photopolymerization of vinyl acetate in ethyl acetate and *n*-hexane solutions were investigated with a view to studying the effect of environment on the course of the reaction. The ethyl acetate solutions allow the polymer to remain extended so that the course of the reaction is similar to that of the pure monomer. With *n*-hexane, precipitation of the polymer takes place with the result that a considerable 'gel' effect is observed. The reaction rate in both cases is proportional to the monomer concentration and to the square root of the light intensity. In ethyl acetate solutions the activation energy is 4·42 kcal./g.mol., which is virtually the same as for the bulk reaction, while in normal hexane the value found is −1·30 kcal./g.mol. Reasons for this have been suggested.

The life time of the active polymer in *n*-hexane solution both before and after the inception of the gel stage has been measured using the sector technique, and the values of the coefficients for the propagation and termination reactions have been computed. It is found that the value of the former is little changed by the change in medium, while that of the termination step falls sharply—the fall being in agreement with the increase in rate. These results confirm present theories on the cause of the gel stage.

INTRODUCTION

The introduction of solvents into polymerization processes leads to a number of effects which depend, to a large extent, on the nature of the solvent used. The range of solvents which can be used in photochemical reactions is naturally limited to those whose ultra-violet absorption is considerably different from that of the compound under investigation.

The effect which is of most interest here is the effect of the solvent on the course of the polymerization reaction. The case of the catalysed polymerization of methyl methacrylate in a variety of solvents has been fully examined (Norrish & Smith 1942). It is found that after a definite amount of conversion the rate of polymerization increases sharply, even although special care is taken to ensure isothermal conditions. This effect occurs at an earlier stage and to a greater extent, the poorer the solvent is with respect to the polymer, giving a maximum effect in the case of precipitants. The cause of this increase in rate is the main interest in this paper. The following suggestions have already been made without the necessary proof. The increase in the rate of reaction may be due to (*a*) increase in the rate of initiation, (*b*) increase in the rate of propagation, (*c*) decrease in the rate of termination. The first idea is easily discounted in view of the fact that the increase in rate is accompanied by an increase in the molecular weight of the polymer formed, which is obviously contrary to the fact that there is an increase in the rate of starting the chains. (The increase in the rate has been termed the 'explosive-like' stage of the reaction by Schulz & Blaschke (1941), but we consider that this term is misleading, since the reaction does not involve the branching of chains, so that throughout the phenomenon will be referred to as the 'gel effect' since its appearance is bound up with the increasing viscosity of a solution, or the actual formation of a gel.) It is also very difficult to explain satisfactorily how an increase in the rate of the propagation step may take place. The decrease in the rate of termination may be readily accounted for, in the case of pure monomer, by the increase in the viscosity, putting a great enough restriction on the motion on the growing molecules to prevent the active ends from making contact and so preventing chain breaking. This would account for the increase both in the rate and the degree of polymerization. With solutions the phenomenon appears to depend on the coiling of the macromolecule in poor solvents, again preventing the contact of the active ends to terminate the growth of the chain. From this theory it is obvious that the rate and degree of polymerization will not increase if termination occurs by the anomalous addition of a monomer molecule. The photopolymerization of vinyl acetate in two solvents—ethyl acetate and *n*-hexane—has been studied with a view to clarifying the position.

The second effect, which is probably not of very great importance here, is that of transfer, whereby the reaction of a growing polymer molecule with a solvent molecule results in the production of a dead polymer and a solvent fragment which is capable of initiating a fresh chain. This may be shown by the following reaction sequence:

$$\begin{array}{cll}
M\rightarrow P_1 & \text{initiation} & \text{I}\\
P_n+M\rightarrow P_{n+1} & \text{propagation} & \text{II}\\
\begin{array}{c} P_n+XS\rightarrow M_n+S- \\ S-+M\rightarrow P_1\end{array} & \text{transfer} & \left\{\begin{array}{l}\text{III}a\\ \text{III}b\end{array}\right.\\
\multicolumn{2}{c}{2S-\rightarrow\text{reaction products}} & \text{IV}
\end{array}$$

in which M is a monomer and XS a solvent molecule, P_n is the active polymer molecule containing n monomeric units and M_n is the corresponding dead polymer. For chain transfer to occur without a decrease in the velocity of polymerization reactions III*a* and III*b* will be fast compared to IV. If this is not the case then the velocity of reaction will fall, and in the extreme case where IV is very much faster than III*b* the solvent will act as a retarder. In any of these cases the degree of polymerization will fall.

For free radical mechanisms it is obvious that the easier the formation of free radicals by the solvent the more likely is transfer to occur. This question has been investigated by Mayo (1943) on the basis of results by Suess, Pilch & Rudorfer (1937) and Schulz & Husemann, (1939).

EXPERIMENTAL

The *n*-hexane used was of spectroscopic purity. Ethyl acetate was purified by fractional distillation, and the fraction boiling, under atmospheric pressure, at 77 to 78° C was retained. The taps of the vacuum apparatus were greased with silicone, thickened with talc, in the case of hexane. Both the solvents and the monomer were thoroughly freed from air by freezing in liquid air and evacuating, as was described in an earlier paper. The experiments were carried out under vacuum in sealed tubes.

The monomer which was thoroughly outgassed was distilled under high vacuum into a graduated tube from which a known volume could be distilled to the reaction tube. The solvent was then distilled in to make a known volume of solution. The reaction was followed dilatometrically using a cathetometer to measure the fall in the level of the meniscus.

ETHYL ACETATE SOLUTIONS

Solutions of the monomer in ethyl acetate were made up, as has been described, having concentrations from 1·5 to 4·5 g.mol./l. of vinyl acetate. Irradiation of the solution with ultra-violet light produced a ready polymerization: the course of the reaction is shown in figure 1. It will be seen that, as in the case of the bulk photopolymerization, the rate of reaction is initially constant. After some time the rate begins to fall off as might be expected as the monomer is used up. It will be noticed also that there is no sign of the inception of the gel stage in the polymerization, so that in this case it may be readily assumed that the macromolecule remains extended in solution, thus placing no very great restriction on the motion of the active ends of the polymer molecules. The steady state concentration of the active molecules must also be very rapidly established and the life time of the active species short, since the reaction ceases immediately on the removal of the source of radiation.

Melville & Gee (1944) have shown that in reactions in which the termination is due to the interaction of two of the growing polymer molecules, the rate of polymerization is proportional to the square root of the light intensity and also to the first power of the monomer concentration. Since the former may also depend on the monomer concentration to some extent, the order of the reaction with respect to the monomer may lie between 1 and 1·5. To check this a variety of solutions were used and the rates of the initial polymerization measured, the results being set out

TABLE 1. EFFECT OF CONCENTRATION OF VINYL ACETATE

rate (R) (g.mol./l./min.)	concentration of monomer (C) (g.mol./l.)	R/C	$R/C^{\frac{1}{2}}$	$R/C^{\frac{3}{2}}$
0·0410	4·5	0·0091	0·0193	0·0043
0·0279	3·0	0·0093	0·0161	0·0053
0·0142	1·5	0·0095	0·0116	0·0077

in table 1. From these figures it is evident that the rate of polymerization is proportional to the first power of the monomer concentration. This fact supports the assumption which was tacitly made in the calculation of the kinetic coefficients of the bulk reaction, and which was reported in a previous communication.

The intensity of the radiation was varied by means of perforated metal screens the transmission of which was known and the rate of polymerization was determined, using, in all cases, solutions containing 1·5 g.mol. of vinyl acetate per litre. The results which are tabulated in table 2 gave a value of 0·51 as a mean for the intensity exponent. Thus the termination process is the same as before, i.e. by the interaction of two of the active polymer molecules.

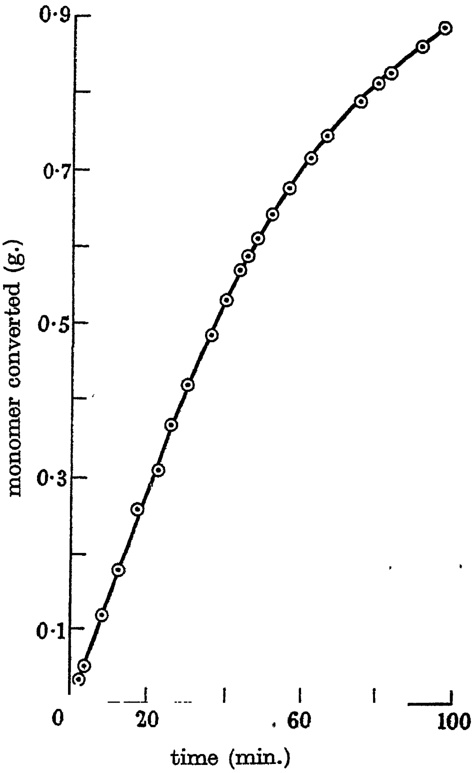


Figure 1. Course of polymerization of vinyl acetate in ethyl acetate solution.

Table 2. Intensity dependence

relative intensity	relative rate	intensity exponent
0·42	0·70	0·50
0·25	0·48	0·52
0·16	0·39	0·52
0·06	0·25	0·49

Earlier it was mentioned that in solution there is the possibility of transfer. Even if this does occur it will not invalidate our conclusions as the following analysis shows. The sequence of reactions will be:

$$\begin{array}{lll}
M \rightarrow P_1 & f(I) & \mathrm{I} \\
P_n + M \rightarrow P_{n+1} & k_{p_n} & \mathrm{II} \\
P_n + XS \rightarrow M_n + S- & k_{r_n} & \mathrm{III}a \\
S- + M \rightarrow P_1 & k_f & \mathrm{III}b \\
2S- \rightarrow ? & k_s & \mathrm{IV} \\
P_n + P_s \rightarrow M_n + M_s & k_{m_n}, k_{m_s} &
\end{array}$$

The stationary state equations will be

$$\frac{d(P_1)}{dt} = f(I) + k_f(M)(S-) - k_{p_1}(P_1)(M) - k_{r_1}(P_1)(XS) - k_{m_1}(P_1)\Sigma k_{m_s}(P_s) = 0,$$

$$\frac{d(P_2)}{dt} = k_{p_1}(P_1)(M) - k_{p_2}(P_2)(M) - k_{r_2}(P_2)(XS) - k_{m_2}(P_2)\Sigma k_{m_s}(P_s) = 0,$$

and, in general,

$$\frac{d(P_n)}{dt} = k_{p_{n-1}}(P_{n-1})(M) - k_{p_n}(P_n)(M) - k_{r_n}(P_n)(XS) - k_{m_n}(P_n)\Sigma k_{m_s}(P_s) = 0,$$

and also the equation which will govern the concentration of the free radicals formed from the solvent in the transfer reaction will be

$$\frac{d(S-)}{dt} = (XS)\Sigma k_{r_s}(P_s) - k_f(M)(S-) - k_s(S-)^2 = 0.$$

Now if transfer is to occur readily then it is obvious that reaction IIIb is to be very much faster than reaction IV, so that the last term of this equation can be omitted. Hence the concentration of the free radicals will be given by

$$(S-) = \{(XS)\Sigma k_{r_s}(P_s)\}/k_f(M).$$

Adding to infinity the stationary state equations given above, then

$$f(I) + k_f(M)(S-) - (XS)\Sigma k_{r_s}(P_s) - \{\Sigma k_{m_s}(P_s)\}^2 = 0.$$

Substituting the value of $(S-)$ reduces this expression to

$$f(I) = \{\Sigma k_{m_s}(P_s)\}^2.$$

Now the rate of reaction will be given by

$$-d(M)/dt = f(I) + k_f(M)(S-) + (M)\Sigma k_{p_s}(P_s).$$

Substituting the value of $(S-)$ and considering the chains to be long when it is possible to neglect the first term, this will give

$$-d(M)/dt = (XS)\Sigma k_{r_s}(P_s) + (M)\Sigma k_{p_s}(P_s).$$

If the usual assumption that the ratio of the coefficients is independent of the molecular size is made, then

$$\delta^{-1}k_{m_s} = k_{p_s} \quad \text{and} \quad \mu^{-1}k_{m_s} = k_{r_s},$$

which will give the rate of reaction as

$$-\frac{d(M)}{dt} = \delta^{-1}(M)\,\Sigma k_{m_s}(P_s) + \mu^{-1}(XS)\,\Sigma k_{m_s}(P_s),$$

and substituting the value of $\Sigma k_{m_s}(P_s)$ in this gives the final result

$$-d(M)/dt = f(I)^{\frac{1}{2}}\{\mu^{-1}(XS) + \delta^{-1}(M)\}.$$

It will, therefore, be seen that, provided the chains are eventually broken by the mutual reaction, the overall rate of reaction is still proportional to the square root of the light intensity.

The rate of polymerization was determined at temperatures ranging from 31 to 63° C, and gave the results shown in table 3.

TABLE 3. THE ENERGY OF ACTIVATION

temperature °K	rate (arbitrary)	E kcal./g.mol.
304·6	0·84	4·45
313·6	1·06	4·55
326·0	1·44	4·28
336·3	1·79	

E = 4·42 kcal./g.mol.

These results indicate that the photopolymerization of vinyl acetate in ethyl acetate solution follows a course which is identical with that obtained in the case of the bulk polymerization of the monomer under the same conditions. In view of the fact that the structure of the solute and solvent are so similar this is not altogether surprising. The results which have been given here tend to indicate that the light responsible for the polymerization is absorbed completely within the reaction system, since otherwise the rate of reaction would have been proportional to some power of the monomer concentration greater than unity, as has been pointed out in a previous paper.

The rates obtained in this case are of the same order as those obtained in the polymerization of the pure monomer, so that it is unlikely that transfer can occur to any great extent. This is not wholly surprising, for the only method whereby transfer could occur would be by the removal of a hydrogen atom from the ethyl acetate which would be a relatively difficult matter. It is well established that compounds containing only carbon, hydrogen and oxygen are not efficient transferring agents. Thus it may be said that transfer, if it occurs at all, does not do so to any appreciable extent.

POLYMERIZATION IN *n*-HEXANE SOLUTION

The course of the photopolymerization of *n*-hexane solutions of vinyl acetate is shown in figure 2. Almost immediately the light is switched on an opalescence is seen in the reaction tube and, as the time of radiation increases, a definite amount of polymer is formed on the walls and does not remain in suspension indefinitely. From figure 2 it will be seen that the rate of polymerization remains constant until a certain amount of polymer has been formed when the rate increases abruptly and continues to increase over a considerable period. The reason for this phenomenon will be discussed in detail later. After a further period the rate begins to fall off as would be expected from the decrease in the concentration of the monomer.

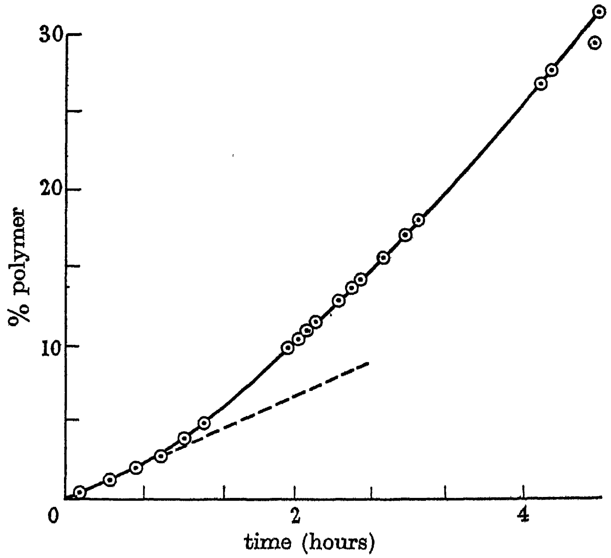


FIGURE 2. Course of polymerization of vinyl acetate in *n*-hexane solution.

TABLE 4. DEPENDENCE OF RATE ON CONCENTRATION (*n*-HEXANE)

concentration of monomer (C) (g.mol./l.)	rate (R)	R/C	$R/C^{\frac{1}{2}}$	$R/C^{\frac{3}{2}} \times 10^2$
6·0	0·0143	0·0023	0·0059	0·098
4·5	0·0102	0·0023	0·0048	0·107
3·0	0·0073	0·0024	0·0042	0·140
1·5	0·0036	0·0024	0·0029	0·196

Concentrations of monomer from 1·5 to 6 g.mol./l. were used, and for purposes of comparison only the initial stages of the reaction were considered, i.e. up to about 5 % conversion of monomer. This removes any ambiguity which might be attributed to the abnormal stage of the reaction. The results of the experiments are set out in table 4. From these figures it is obvious that the reaction is again of first order with respect to the monomer concentration, at least, during the initial stages of the reaction.

As before, the variation of the rate with the light intensity was measured by changing the intensity by means of perforated screens of known transmission, since the inception and extent of the gel stage of the reaction depends on the light intensity, as will be seen from figure 3, in which the upper curve is for high intensity. Here it is evident that the lowering of the molecular weight due to the high rate of initiation has effected the inception of the gel stage to some extent. This will be discussed in greater detail later. The results of the experiments are recorded in table 5. Again it will be seen that the rate of polymerization is proportional to the square root of the intensity of the absorbed radiation, since the intensity exponent is 0·50. This is not particularly surprising in view of the inert nature of the solvent.

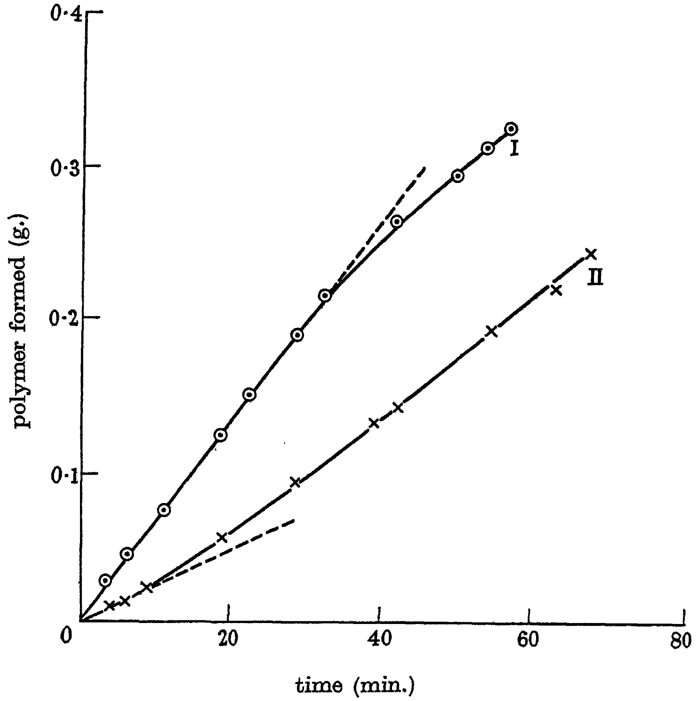


Figure 3. The effect of light intensity on the gel stage. Curve I, very high intensity. Curve II, low intensity.

Table 5. Intensity dependence in n-hexane solution

relative intensity	relative rate	intensity exponent
0·42	0·69	0·48
0·25	0·51	0·50
0·16	0·41	0·49
0·06	0·25	0·49

The variation of the rate of reaction with temperature was carried out over a considerable range of temperature and again only in the initial stages of the reaction. This is again absolutely necessary, since it is certain, as will be shown, that the abnormal stage in the reaction does depend on the temperature. The results are shown in table 6, while the usual logarithmic plot of the results gives the energy of activation as −1·30 kcal./g.mol.

TABLE 6. THE ENERGY OF ACTIVATION IN *n*-HEXANE SOLUTION

temperature °K	rate (arbitrary units)	log (relative rate)
302·6	1·61	0·086
309·0	1·53	0·065
320·0	1·43	0·033
333·1	1·32	0·000

$E = -1{\cdot}30$ kcal./g.mol.

The negative value of the energy of activation is of considerable interest. In gas-phase polymerizations it is almost universally true that the overall energy of activation is negative. This case forms a parallel to the vapour-phase case in that the polymer, when formed, is precipitated, and hence the reaction must to a certain extent cease to be occurring wholly in the one phase. The solution case is relatively understandable, since the overall energy of activation must be made up of a large number of $\exp(E/RT)$ terms to cover the precipitation of the polymer, the adsorption of the monomer on the polymer, etc., as well as the usual terms for initiation, propagation and termination.

The molecular weight of the polymer formed in the initial stages of the reaction (up to 5 % polymerization) and for the almost completed reaction (90 %) were found viscosimetrically using the value of K_m in the Staudinger equation given by McDowell & Kenyon (1940). These results are shown in table 7. This method of determination of the molecular weights was adopted, since there was no need to know the value of the number average molecular weight, but merely to check that the effect was of the same nature as that found by Norrish & Smith. Since the low molecular weight polymer is still present in the final polymer, it is obvious that there must have been a considerable increase in the molecular weight of the polymer formed in the later stages of the reaction. If the Staudinger equation holds then it is easy to show that

TABLE 7. MOLECULAR WEIGHT OF POLYMER FROM *n*-HEXANE SOLUTION

stage of reaction	intrinsic viscosity (in benzene)	molecular weight
before gel effect	0·297	11,400
after completion of reaction	0·405	15,500

$$\eta_2 = \frac{\eta_t c_t - \eta_1 c_1}{c_2},$$

where η_1 and η_2 are the viscosities of the two fractions of concentrations c_1 and c_2 respectively. If the concentration of the low molecular weight material is taken as 5 %, it is possible to obtain a somewhat crude estimate of the molecular weight of the final polymer formed. The value which is obtained by substituting the appropriate figures in the above equation is about 20,000. This shows a considerable increase over the value obtained in the earlier stages of the reaction. Hence it may be concluded that this is a real gel polymerization.

LIFE TIME OF THE ACTIVE PARTICLE IN *n*-HEXANE

To obtain the life time of the active polymer in *n*-hexane the sector method, described in a previous paper, was used. The procedure had to be simplified somewhat on account of the complex nature of the reaction. It is not possible, as previously, to vary the speed of the sector throughout a single run, but, as was pointed out, this is not necessary, nor, in this case, is it particularly desirable. It will obviously be necessary to conduct complete runs with the sector running at a constant speed throughout, in order to arrive at results which will explain the curves obtained in the gel stage of the polymerization. The sector was run at a speed to give a flash of light of 7·56 msec. duration—a value which is considerably less than the life time of the active particle—and at a speed corresponding to 140 msec., when a rate lower than the former was obtained. By interpolation on the graph for the theoretical function, a value for the absolute life time of the particle can be found.

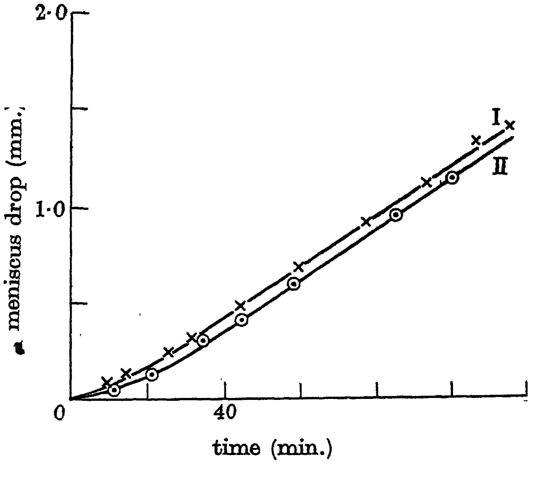


FIGURE 4. Sector experiments. Curve I, flash duration = 7·56 msec. Curve II, flash duration = 140 msec.

Figure 4 shows the shape of the curves obtained. On account of the low light intensity the inception of the gel stage was naturally very rapid, but the curves are clearly enough defined to make the necessary calculations possible. It will be seen that in curve II, in which the sector was rotating slowly, the initial rate of polymerization is much less than in curve I, but eventually the two curves run parallel to one another. This obviously means that the value of the life time of the active polymer changes from somewhat less than 140 msec. to a value greater than this.

From the analysis given in the previous paper it is possible from these results to determine the life time of the active species and hence to compute the values of the kinetic coefficients of the reaction. The change in the initial rates of polymerization corresponds to a value of 81 msec. for the life time, and we estimate the life time at some stage in the gel reaction to be 170 msec. This is reasonable, since it is known that the life time will increase throughout this stage from the fact that it appears from the results to do so continuously. This means that the rates in the initial and final stages of the reaction should be in a ratio of the order of 1:4. This qualitatively fits the facts, but it still remains to calculate the coefficients of the propagation and termination reactions, k_p and k_t respectively. The rate of polymerization which can be deduced from the contraction of the monomer during polymerization is $5{\cdot}0 \times 10^{-6}$ g.mol. l.$^{-1}$ sec.$^{-1}$, and the rate of initiation is $6{\cdot}0 \times 10^{-8}$ g.mol. l.$^{-1}$ sec.$^{-1}$. Applying Melville & Jones's (1940) approximate equation

$$\frac{d(P)}{dt} = \frac{(P)}{\tau},$$

where (P) is the concentration of the active particles and τ is the life time, then

$$(P) = 4{\cdot}8 \times 10^{-9}\ \text{g.mol./l.}$$

The overall rate of polymerization will be given by the equation

$$-d(M)/dt = k_p(P)(M),$$

which by substitution of the appropriate numbers gives a value of $7{\cdot}0 \times 10^2$ for k_p. In the same manner the value of k_t is found to be $2{\cdot}6 \times 10^9$ l.g.mol.$^{-1}$ sec.$^{-1}$.

Proceeding in the same way the values of the coefficients after the inception of the gel stage can be calculated. These results, together with those already published, are shown in table 8.

TABLE 8. KINETIC COEFFICIENTS FOR VINYL ACETATE POLYMERIZATION

	bulk reaction	normal solution	abnormal*
overall rate (g.mol./l./sec.)	$5{\cdot}5 \times 10^{-5}$	$5{\cdot}0 \times 10^{-6}$	$1{\cdot}1 \times 10^{-5}$
initiation rate (g.mol./l./sec.)	$1{\cdot}6 \times 10^{-7}$	$6{\cdot}0 \times 10^{-8}$	$6{\cdot}0 \times 10^{-8}$
k_p (l.g.mol.$^{-1}$ sec.$^{-1}$)	$6{\cdot}7 \times 10^2$	$7{\cdot}0 \times 10^2$	$6{\cdot}8 \times 10^2$
k_t (l.g.mol.$^{-1}$ sec.$^{-1}$)	$2{\cdot}5 \times 10^9$	$2{\cdot}6 \times 10^9$	$5{\cdot}0 \times 10^8$
τ (msec.)	51	81	170

* After 5 % polymerization.

GEL POLYMERIZATION PROCESSES

In the catalytic polymerization of methyl methacrylate it was found that, even if the reaction was strictly isothermal, after a certain amount of polymerization the rate of conversion of the monomer increased rapidly. At the same time there was an increase in the degree of polymerization. As was explained before, because of the formal similarity between the reaction and the usual type of explosion reaction this has been termed the explosion stage, but to avoid unnecessary confusion in ideas, we have adopted the term gel polymerization. The increase in the degree of polymerization indicated quite clearly that the increase in the rate could not be due to an increase in the rate of initiation, since it is known that this is accompanied by a decrease in the molecular weight. Also it is not easy to conceive how it is possible for the rate of the propagation to increase. Hence it appears to be possible that the reaction step which is responsible for the increased rate is the termination process.

The gel stage can be brought about very readily by the introduction of solvents. Norrish & Smith (1942) investigated this effect using a range of solvents which were both good and bad solvents for the polymer. They found that the inception of the gel stage was earlier the poorer the solvent, and was most marked in the case in which the solvent acted as a precipitant for the polymer. Thus in solvents such as heptane the increase in rate set in very early, while in more efficient solvents more than 50 % conversion was required. The investigation was carried a stage further by Trommsdorff (1944), who dissolved in pure methyl methacrylate quantities of cellulose tripropionate. This had the effect of increasing the viscosity of the monomer and caused a much earlier appearance of the gel stage of the polymerization.

The inception and extent of the gel stage in the polymerization was found to depend on (*a*) the concentration of the catalyst and hence the rate of starting of chains and the degree of polymerization, and (*b*) the temperature. In the case of vinyl acetate photopolymerization in *n*-hexane solution, runs at different intensities showed that the higher the light intensity the more delayed was the inception of the gel stage, as will be seen from figure 3. The raising of the temperature had the same effect, and at sufficiently high temperatures the increased rate stage of the reaction could be eliminated altogether (figure 5). The increasing of the intensity of radiation (or of the catalyst concentration) will lead to a decrease in the degree of polymerization, and so, in the case of the polymerization of the pure monomer, lead to a decrease in the viscosity of the reaction medium as would an increase in the temperature. Hence it is obvious that the viscosity of the medium must play an important part in the course of the reaction. In the case of solvents it is evident that the formation of a gel or the coiling of the molecule is the feature which will be of most importance.

The current theory due to Trommsdorff, Schulz and others is that the increase in the viscosity of the medium either in solution or in pure monomer will put considerable restriction on the motion of the growing polymer molecules. The result of this restriction is that the active ends are unable to make contact and so break the reaction chains. The rate of the diffusion of the monomer molecules to the active ends will not be seriously affected by the change in the medium, so that the number of collisions between the monomer and the active end will remain virtually unchanged. Thus the propagation reaction is not seriously affected by the change in the medium. On the other hand, the lowering of the number of collisions between the active ends will reduce the termination rate, so that both the increase in the rate and the degree of polymerization can be accounted for.

The degree to which the gel polymerization will be shown by monomers will depend on how sensitive the polymer is to the restriction in its motion. If the polymer is not abnormally kinked the active end will be subjected to severe restriction in motion even if the actual restriction is applied at a relatively remote point. With kinked polymers the effect will not be so great. This probably accounts for the fact that the phenomenon is so clearly marked in the case of the methyl methacrylate, in which it is known that the polymer is not severely kinked.

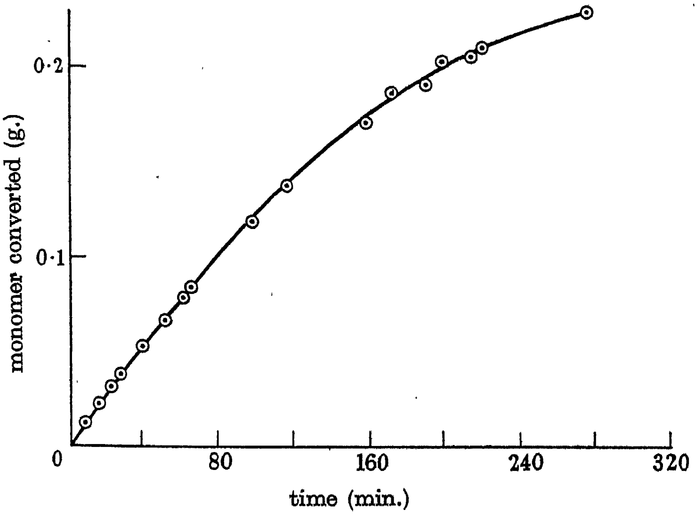


FIGURE 5. Effect of temperature on the gel stage. Temperature = 65° C.

In poor polymer solvents the polymer must be severely coiled so that it is most likely that a similar state of affairs will exist. The coiling of the molecule will prevent the active ends from coming into contact with the normal frequency, so that the rate of chain breaking will be reduced. From the results which were given in the previous section it is evident that the above theory is tenable. The value of the velocity coefficient for the propagation step is not changed to any measurable extent during the abnormal part of the reaction and is not very much different from the value obtained in the case of the bulk polymerization. On the other hand, the value of the termination coefficient falls considerably—in our case by a factor of about 5. This means that the rate of polymerization should have increased by a factor of the order of 2·3, since the rate is inversely proportional to the square root of the termination velocity coefficient. This is also verified from the results which have been obtained. Trommsdorff, in his paper, applies this theory to explain the reason for the very high rates and degrees of polymerization encountered in the case of emulsion polymerization processes.

The value of the temperature coefficient in the reaction can be explained by the fact that as the temperature is increased the velocity of the reaction may also increase, but this is offset by the fact that the stiffness of the gel or the viscosity of the solution is lowered, thereby decreasing the velocity, which will be controlled primarily by the stiffness of the gel. This will mean that the reaction will not be so temperature dependent as in the case of the bulk polymerization, and may, indeed, lead to a negative value of the temperature coefficient. The idea of the formation of a new phase having an effect on the course of the reaction is not new. As has been suggested, this must definitely happen in all vapour-phase polymerization processes, so that is is not wholly surprising that in these reactions the value of the overall energy of activation is negative.

One of us (G.M.B.) is indebted to Distillers Co. Ltd. for financial assistance (1945–6).

REFERENCES

Jones, T. T. & Melville, H. W. 1940 *Proc. Roy. Soc.* A, **175**, 392.
McDowell, W. H. & Kenyon, W. O. 1940 *J. Amer. Chem. Soc.* **62**, 415.
Mayo, F. R. 1943 *J. Amer. Chem. Soc.* **65**, 2324.
Melville, H. W. & Gee, G. 1944 *Trans. Faraday Soc.* **40**, 240.
Norrish, R. G. W. & Smith, R. R. 1942 *Nature*, **150**, 336.
Schulz, G. V. & Blaschke, P. 1941 *Z. phys. Chem.* **50**, 305.
Schulz, G. V. & Husemann, E. 1939 *Z. phys. Chem.* B, **43**, 385.
Suess, H., Pilch, K. & Rudorfer, H. 1937 *Z. phys. Chem.* A, **179**, 361.
Trommsdorff, E. 1944 *Colloquium on high polymers*, Freiburg. See *B.I.O.S. Rep.* no. 363, item no. 22.

The dehydration of ammonium, potassium and some mixed alums

By G. P. Acock, W. E. Garner, F.R.S., J. Milsted and H. J. Willavoys

(*Received* 2 *August* 1946)

[Plates 12 to 14]

The shapes of the dehydration nuclei and their rates of growth have been investigated in hard vacuum and in the presence of water vapour. The activation energies and *B*-factors for the interface reaction in hard vacuum are those normally expected for solid reactions occurring at room temperature. The thermodynamic heat of dissociation is within ±0·5 kcal. of the heat of activation, indicating a very close coupling between the process of evaporation of water and the growth of the new solid phases. The main process occurs layer by layer along the 100 planes of the hydrate. There is evidence of the production of a lattice defect of water in the hydrates before nuclei are formed.

Water vapour slightly accelerates the rate of reaction at very low water pressures, but the main effect at higher pressures is a retardation which is probably due to the blocking of active areas in the interface by adsorbed molecules of water.

In mixed potassium and chromium alums, the layer growth is inhibited and dehydration occurs ahead of the normal interface, as if the hydrates were behaving as zeolites. The reaction can occur in several stages and very complex phenomena are observed.

Explanations are given of the abnormal reactions occurring with some hydrates, in which there is a breakdown of the close coupling between the dehydration process and the formation of the new solid phases.

Introduction

During the dehydration of salt hydrates in a vacuum or over a dehydrating agent, nuclei are formed of a new solid phase on the surfaces of the crystals. The majority of recent investigations have been concerned with the study of the rate and of the mechanisms of the reactions occurring at the interface between the new solid phase and the original hydrate. The reaction is followed by measuring the change in weight of the crystal during dehydration, or by observation of the rate of movement of the interface. The activation energy for the evaporation of water from the interface and the *B*-factor in the equation $k = Be^{-E/RT}$ have been obtained for a number of hydrates.

In some of the examples the values of *B* and *E* are of the order normally expected for solid reactions occurring at room temperature. For copper sulphate pentahydrate (Topley 1932*a*) and nickel sulphate pentahydrate (Garner & Southon 1935), the *B*-factor lies in the region 10^{11} and 10^{12} mol./cm.^{2}/sec. In the case of the former, $E = 18{\cdot}25$ kcal., and is of a similar order of magnitude to the thermodynamic heat of dissociation, viz. 12·6 kcal. Carbonates dissociate similarly to hydrates, and in a number of examples the heat of dissociation is approximately the same as the heat of activation for the loss of carbon dioxide (Zawadzki & Bretsznajder 1938). For silver carbonate the same is approximately true (Spencer & Topley 1929) and the *B*-factor is normal. There would appear, therefore, to be one class of dissociations for which the B-factor is of normal magnitude ($\sim 10^{13}$) and the heat of dissociation approximately the same as the heat of activation.

On the other hand, calcium sulphate hexahydrate (Topley & Hume 1928) gives a high activation energy, $E = 38$ to 40 kcal. and a B-factor of 10^{36}, and for chrome alum (Cooper & Garner 1940) $E \doteqdot 31$ kcal. as compared with 11 kcal. for the heat of dissociation and a B-factor of 10^{25}. Potassium hydrogen oxalate (Hume & Colvin 1929) also gives an abnormally high value for the activation energy, $E = 40$ kcal. No satisfactory explanation has yet been given of this abnormal behaviour.

If the crystals have been carefully grown and the surfaces are free from scratches, there is normally an induction period before nuclei make their appearance. In extreme cases, hydrates have been kept for several days in a high vacuum in the neighbourhood of room temperature without visible change occurring. This indicates that the activation energy for the removal of single water molecules from the surface of the crystal must be very high, but so far there has been no accurate estimate of its magnitude for any hydrate. This is in marked contrast to the removal of water from the interface, where the activation energy may be as low as the heat of dissociation.

The nuclei frequently adopt striking geometrical patterns with sharply defined contours. An extended investigation into the shapes of nuclei on the known faces of copper sulphate pentahydrate (Garner & Pike 1937) showed that the shapes on different crystal faces are often markedly different, and the work indicated that growth along certain crystal axes was preferred. For nickel sulphate heptahydrate linear nuclei were obtained on the surface, which spread into the interior of the crystal along a preferred plane. On the other hand, in the case of chrome alum (Cooper & Garner 1940) hemispherical nuclei are obtained, where the rate of growth of the nuclei is the same for all crystal directions.

A factor which complicates the study of the rates of growth of nuclei is that of the impedance to the escape of water, due to a fine state of division of the solid product. This may lead to a decrease in the rate of growth as the thickness of the layer on product increases. Impedance is marked in the case of copper sulphate pentahydrate (Topley 1932*a*), where the product has been shown to be amorphous when examined by X-rays (Kohlschütter & Nitschmann 1931). Manganous sulphate bihydrate also gives an amorphous product (Smith & Topley 1935; Volmer & Seidel 1937), which crystallizes in the presence of water vapour, and in this case a very abnormal rate of growth curve is obtained in the presence of water vapour, as shown when the rate of growth is plotted against pressure of water vapour. This is not, however, the general rule, for in other cases, notably the alums, there is no appreciable impedance to the escape of water vapour due to increasing thickness of product. The phenomena which are observed must depend on the compactness of the layer of solid product. If this is free from cracks or capillary passages of large size, then some impedance might be expected.

In the present investigation a number of alums has been examined, with a view to a study of the interrelationship of the various factors, e.g. B-factor, activation energy, impedance and shape of nuclei. It was hoped thereby to fill in certain gaps in the body of information on nuclei formation and growth, and to attempt a broad generalization of the nature of the mechanism of the dehydration processes. The dehydration of potassium and ammonium alums and solid solutions of potassium aluminium and potassium chromium alums have been studied, using the technique developed by Cooper & Garner (1940). Measurements have been made of the rate of growth of nuclei, and of the rate of penetration of the interface into the interior of the solid for a range of temperatures. The thermodynamic heat of dissociation has been measured for ammonium alum. The shape of the nuclei has been investigated on the different faces of the alum crystals, and in this connexion some work has been done on rubidium and caesium alums, and on mixed alums containing common alum and chrome alum in solid solution. The effect of variation of water pressure on the rate of reaction has also been examined.

EXPERIMENTAL

Measurements of the rates of growth and observations on the shapes of nuclei have been made in the apparatus previously described in the paper on chrome alum (Cooper & Garner 1940). The crystals were grown in a dust-free crystallization chamber, dehydration of the solutions being effected over solid sodium hydroxide. They were grown slowly and only crystals with perfect faces were used. They were drained from excess of solution and dried by the method previously described and stored in desiccators over the powdered alum. Crystals of 1 to 2 cm. across were employed.

Shapes of nuclei formed in vacuum

The nuclei formed on the 111 face of the crystals are roughly hexagonal in shape, with marked diagonal cracks. Typical nuclei for the series ammonium, potassium, rubidium and caesium alums are shown in figures 1 and 3, plates 12 and 13. For ammonium, potassium and rubidium alum, the alternate sides of the hexagon are concave and the other three sides give very nearly a straight edge, but with an occasional tendency to concavity and with an overhanging haze. The haze, under higher power, gives the appearance of being composed of very fine lines. The nuclei for caesium alum are somewhat more complex in structure, since three of the faces of the hexagon are poorly developed (figure 3*b*). The nuclei all possess trigonal symmetry and are aligned with their straight edges parallel to the edges of the 111 face; hence they form an approximate reproduction of the octahedral face of the crystal.

On the cubic face, 100, the nuclei are square with diagonal cracks, giving the appearance of a square pyramid (figure 1*c*, *d*). The sides of the square are parallel to the intersection of the cubic face with the octahedral face. On the dodecahedral face, 011, the nuclei are approximately rhombic with two of the corners rounded (figure 1*e*,*f*). The shapes of the nuclei in relation to the geometry of the crystal are shown in figure 2.

The cracks in the nuclei possess some significance, for their formation indicates a regular shrinkage of the product, due to recrystallization. This can be readily interpreted if the microcrystals of the product were regularly arranged with respect to the edges of the nuclei. This is what would be expected to occur if the dehydration takes place preferentially along one set of planes (100) of the hydrate, and the product was built up along these planes. A mass of similarly oriented crystals would be produced which, on recrystallization, would contract in a regular manner.

For ammonium alum, slow-growing spherical nuclei were also formed, especially over the lower temperature range. In this type of nucleus there are no cracks. The microcrystals of the product are therefore less regularly arranged and when recrystallization occurs there is no shrinkage of the mass as a whole, the changes being accommodated between the grains. It is significant that the spherical type of growth might be expected to yield microcrystals arranged with their axes pointing in all directions.

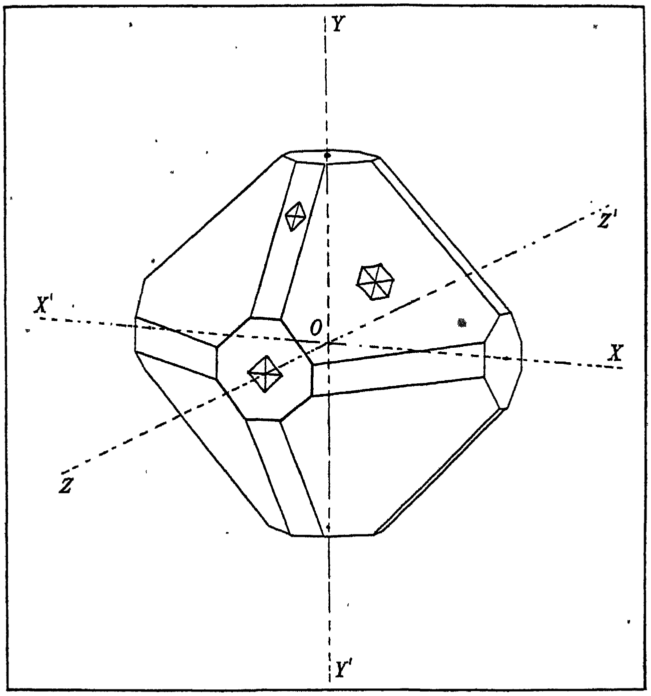


FIGURE 2. Shape of nuclei on faces.

Effect of water vapour on shapes of nuclei

As the water vapour pressure over the crystal increases, the nuclei gradually lose their hexagonal shape and finally become very irregular on the periphery (figure 3 *d*, *e* and *f*, plate 13). At the same time the cracks disappear and the product becomes visibly crystalline in character, indicating that water vapour facilitates the growth of the microcrystals, linking them together irregularly and thereby preventing shrinkage. In the presence of water vapour there is no longer a marked preference for growth along the 100 planes. In figure 3*c*, plate 13, a photograph is shown of the nuclei on iron ammonium alum, produced in hard vacuum, which resembles those of other alums when grown in the presence of water vapour.

Nuclei on the surfaces of mixed alums

A study was made of the nuclei formed on five solid solutions of potassium alum and chrome alum, and typical photographs are given in figure 4, plate 14. It was observed that as the percentage of chromium increases, nucleation becomes more difficult, especially over the lower temperature range (15 to 30° C), and that at the same time there is a marked increase in the temperature coefficient for nuclei formation. The hexagonal structure and the radial cracks of the nuclei on the 111 face are absent and spherical nuclei with concentric cracks are obtained. The growth of these nuclei is discontinuous, wisps of material being developed ahead of the interface at certain stages during the growth. These isolated regions grow both forwards and backwards but never quite join up with the original interface, and they give rise ultimately to what is apparently a concentric crack. Typical nuclei for 30, 40 and 50 % chromium are given in figure 4*a*, *b* and *c*, plate 14, and it will be seen that they all have the same character. At 68 % chromium, the nuclei become very complex and their character depends on the temperature of dehydration. At 35 to 40° C spherical nuclei are formed which resemble those described above in their method of growth, but a mosaic structure is developed in the nuclei (figure 4*d*). At 30° C and below a new type of nuclei is obtained, whether the nucleation is started above 30° C or not. This is a 'cluster' type of nucleus which is sometimes set in a paler matrix. There are apparently two separate processes of dehydration, the interrelationship of which has not been fully worked out; there is a rapid spread of dehydration process with a very faint outline and in this whiter nuclei are subsequently formed (figure 4*e*). Sometimes the faint dehydration is not seen and nuclei similar to those in figure 4*f* are obtained. In this series are also obtained nuclei which show evidence of Liesegang phenomena occurring in the solid product (figure 4*g*).

When the percentage of chromium is still greater (86 %), the nuclei resemble those obtained with chrome alum, giving slow and fast growing spherical nuclei. There is, however, evidence of the presence of a mosaic structure in the faster growing nuclei (figure 4*h*).

It would appear that the replacement of aluminium by chromium leads to the disappearance of the processes of dehydration occurring along crystal planes and to the production of spherical nuclei in which growth has certain discontinuous features. The nuclei show concentric cracks or, in certain cases, a mosaic structure. Dehydration occurs ahead of the normal interface, possibly by the removal of water without lattice collapse, as in zeolites, and then subsequently nucleation occurs in the partially dehydrated material.

Rates of growth

The rates of growth were measured in vacuum and also in the presence of water vapour at pressures lower than that of the dissociation pressure. In vacuum there is an initial induction period before nuclei appear, which decreases with increase in temperature, and for ammonium alum the length of the induction period is 100 min. at 15° C and 3 to 4 min. at 40° C. For mixtures of aluminium and chromium alums (Cr = 68 %) there are no nuclei produced at 25° C in 24 hr., but at 35° C nuclei are formed within a few minutes. Photographs of the nuclei were taken with a Winkel-Zeiss microscope on a Kodak super X-X panchromatic film at intervals during an experiment. When the nuclei were hexagonal or square in shape, they were measured on the film along the diagonals, the average lengths of the diagonals being taken. The plot of the diameter of a nucleus against time practically always showed an accelerating rate of growth when the nuclei were small (figure 5), the curve becoming linear and remaining so during the remainder of the growth. The initial rate of growth is approximately double the final value, and the difference is too great to be accounted for by the cooling of the crystal which occurs as the dehydration proceeds. Also, the abruptness of the change in rate is not in accord with the hypothesis that it is due to the self-cooling of the crystal. It was observed in the case of the alums that nucleation ceased after a certain number of nuclei had been formed.

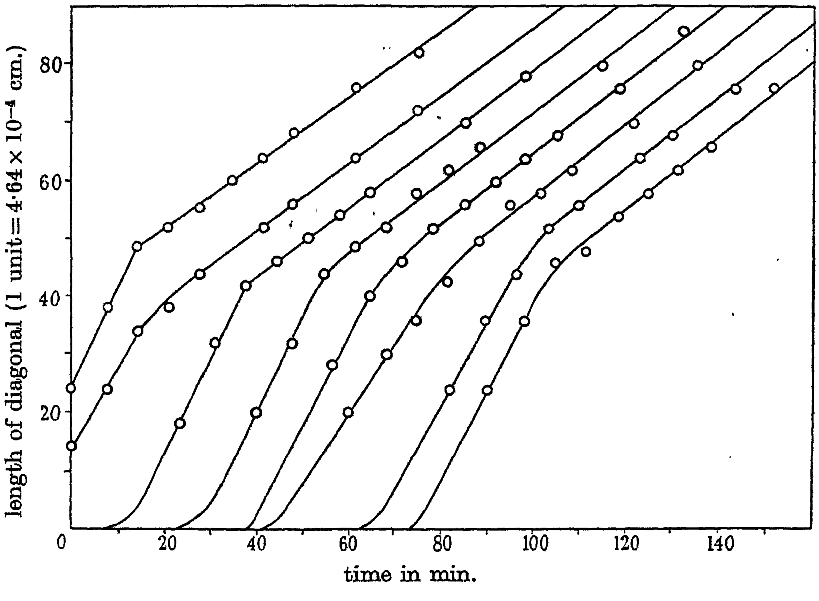


FIGURE 5. Growth of nuclei in chrome alum.

The temperatures of dehydration ranged from 15 to 40° C, and experiments with several crystals from a number of batches were made at each temperature. The average values for the rates of growth of nuclei for ammonium alum on the 111, 100 and 011 faces are summarized in table 1 and for potassium and rubidium alums in table 2. The temperatures are uncorrected for the self-cooling of the crystal during dehydration, but the experimentally measured self-cooling is given (see later).

TABLE 1. RATE OF NUCLEI GROWTH FOR AMMONIUM ALUM IN VACUUM

face of crystal	shape of nuclei	temp. (°C)	self-cooling (°C)	rate of growth (cm./min. × 10^{-4})
111	hexagonal	20	1·10	1·42
		25	1·80	1·95
		30	2·75	2·83
		35	4·15	4·11
		40	6·20	5·36
111	spherical	15	0·55	0·68
		20	1·10	0·97
		25	1·80	1·52
		35	4·15	2·95
		40	6·20	3·90
011	rhombs	25	1·80	1·39
100	square	40	6·20	5·10

TABLE 2. RATE OF NUCLEI GROWTH FOR POTASSIUM AND RUBIDIUM ALUMS IN VACUUM

face of crystal	shape of nuclei	temp. (°C)	self-cooling (°C)	rate of growth (cm./min. × 10^{-4})
		potassium alum		
111	hexagonal	15	—	0·714
		20	1·20	1·121
		25	2·50	1·75
		30	3·55	2·44
		35	4·50	3·69
		40	5·15	5·03
110	squares	40	5·15	4·98
		rubidium alum		
111	hexagonal	25	—	1·37
		28·1	—	1·65

It will be noted that there is a decline in the rate of growth of the nuclei on the 111 face in the order NH_3, K and Rb. From the photographs in figure 1, plate 12 it will be seen that the decrease is more marked in the direction of the concave edges of the nuclei than along the diagonals. The atomic weight of the monovalent ion is obviously a factor influencing the rate of dehydration. For both ammonium and potassium alums the rate of growth along the diagonal of the 111 face is practically identical with that along the diagonal on the 100 face. The growth for the 011 face is definitely slower than the rate on the 111 face.

The results in water vapour were obtained by first forming nuclei in hard vacuum and then continuing the growth in a constant pressure of water vapour. The change-over in rate of growth on introducing water vapour was very rapid. The measured rates of growth are given in tables 3 and 4, and the results, after correction for self-cooling, are plotted in figure 6, together with the earlier results for chrome alum, and it will be seen from the maximum on the curve that there is some catalysis of the rate of dehydration by water vapour at low pressures for potassium alum at 30° C but that the effect is less than for chrome alum. There is a doubtful trace of catalytic effect for ammonium alum at 30° C, but there is no appreciable influence with this hydrate at 40° C. The self-cooling of the crystals was measured and the rates of growth have been corrected on the assumption that the temperature coefficient of the rate of growth is not affected by water vapour.

TABLE 3. RATE OF NUCLEI GROWTH FOR THE 111 FACE OF AMMONIUM ALUM IN WATER VAPOUR

temp. °C uncorr.	vapour pressure (mm.)	rate of growth (cm./min. $\times 10^{-4}$)	temp. °C uncorr.	vapour pressure (mm.)	rate of growth (cm./min. $\times 10^{-4}$)
40	0·00	5·36	30	0·00	2·83
	0·035	4·67		0·014	2·80
	0·313	2·85		0·030	2·19
	0·351	2·41		0·054	2·14
	0·740	1·75		0·168	1·36
	2·130	0·76		0·476	0·66
				1·950	no growth

TABLE 4. RATE OF NUCLEI GROWTH ON THE 111 FACE OF POTASSIUM ALUM IN THE PRESENCE OF WATER VAPOUR

temp. (°C)	vapour pressure (mm.)	rate of growth (cm./min. $\times 10^{-4}$)
30	0·0	2·44
	0·0296	2·47
	0·0966	2·17
	0·286	1·11
	0·476	0·72
	0·776	0·142

TABLE 5. RATE OF GROWTH ON 111 FACE

percentage of chrome alum	rate of growth on 111 face (cm./min. $\times 10^{-4}$)					
	15° C	20° C	25° C	30° C	35° C	40° C
0·0	0·714	1·121	1·75	2·44	3·69	5·03
30·23	—	0·352	0·56	—	—	—
39·75	—	—	0·437	0·690	1·12	—
50·40	—	—	—	—	0·63	1·20
67·98	—	—	(3·30)	(2·40)	(4·93)	(8·20)
85·8	—	(0·927)	(2·2–4·5)	(6·2–7·7)	—	—
100·0	1·10	2·03	3·68	5·27	—	—

Rate of growth in mixed chromium-aluminium alums. The rate of growth of the nuclei is slightly depressed as the percentage of chromium increases up to 50 %. At higher percentages of chromium the formation of nuclei is complex in character, and variable results for the rates are obtained due to the complex interrelationships between two processes of dehydration. In this region 'mosaic' and 'cluster' nuclei are obtained.

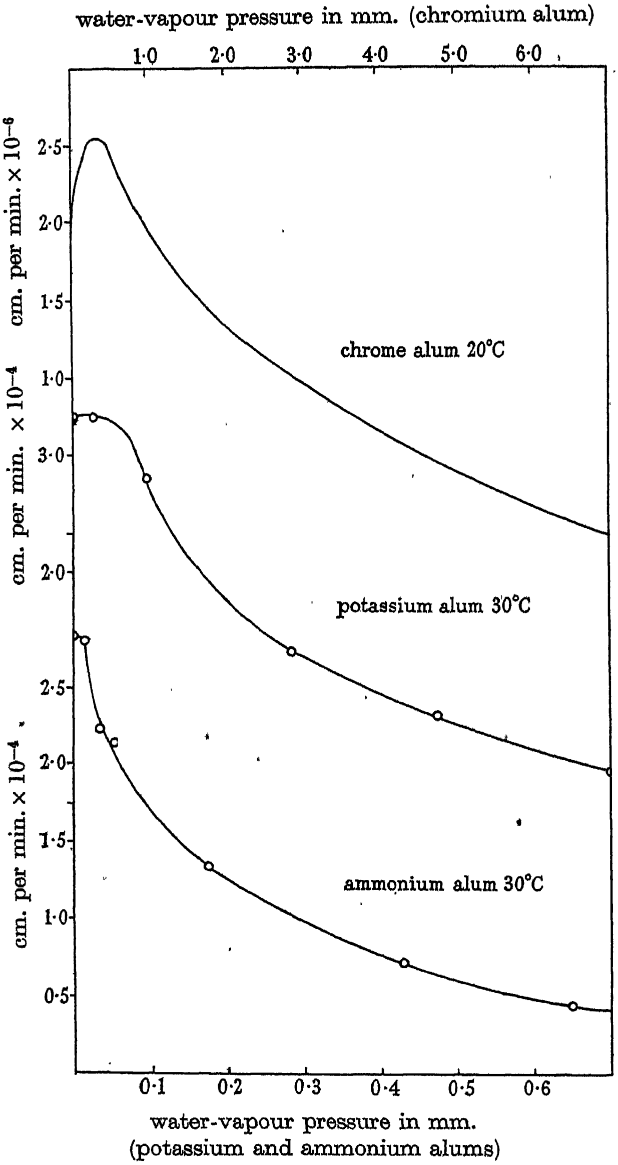


FIGURE 6. Rates of growth of nuclei in water vapour.

Effect of water on the rate of dehydration

In a hard vacuum, the quantity of water adsorbed in the interface will be small, especially where the interface intersects the surface of the crystal. It is shown later that there is no change in the impedance to the escape of water as the thickness of product increases, so that the quantity of water adsorbed in the interface is probably uniform over the whole area. Assuming that the adsorption in a hard vacuum is practically zero and that water retards the dehydration process by occupying places in the interface into which water molecules are released from the hydrate lattice, then it might be expected that the rate of dehydration in the presence of water vapour would be proportional to the places unoccupied by adsorbed water molecules. The fraction of places unoccupied can be calculated from the Langmuir adsorption isotherm, for which $P_{H_2O}/Q = k.P_{H_2O} + \text{const.}$, where Q is the fraction of the places in the interface occupied by water. When $Q = 1$, then the rate $r_p = 0$, and when $Q = 0$ then $r_p = r_0$ and for intermediate values of Q, $r_0 - r_p = kQ$. A plot of $P_{H_2O}/(r_0 - r_p)$ against P_{H_2O} gives a straight line for the two series of experiments with ammonium alum (figure 7), which is in agreement with the proposed mechanism. A similar treatment has been given by Topley (1932*b*) and Smith & Topley (1931) to the results obtained by them for the dehydration of copper sulphate pentahydrate in the presence of water vapour, and they find that there is a linear relationship between the rate of dehydration and the amount of water adsorbed in the interface (see figure 7).

Unfortunately, this treatment cannot be applied to the results obtained with potassium and chrome alums on account of the maxima on the rate curves. The maxima are due to the acceleration of the rate of dehydration at very low water vapour pressures. It is well known that water catalyzes the crystallization of the solid products formed in dehydration processes so that the maxima on the curves for potassium and chrome alums might be ascribed to impedance of the flow of water vapour caused by a transition layer of amorphous solid, which crystallizes in the presence of water vapour.

Rate of penetration of the interface into the crystal

This was measured on a McBain-Bakr quartz spring balance with partially dehydrated crystals which were covered with the product of the reaction over the whole surface. The technique has been described previously. For chrome alum, which gives hemispherical nuclei, it has been shown that the rate of growth along the surface was equal to that of penetration into the crystal, and that the rate of penetration was independent of the thickness of the solid products. From this it was deduced that there was no change in the impedance to the escape of water as the layer of product became thicker and that if any impedance was present it was limited to a transition layer. In the present experiments, the rate of penetration was also found to be independent of the thickness until the reaction was nearly complete, which shows that the type of impedance found with copper sulphate pentahydrate was absent. Twenty out of the original twenty-four molecules of water on R_2SO_4, $Al_2(SO_4)_3 . 24H_2O$ are lost, the final product holding on firmly to the residue of four molecules. There was a tendency for slightly greater losses at the higher temperatures (see tables). The rates of penetration are given in tables 6 and 7 and compared with the radial growth of the nuclei over the surface.

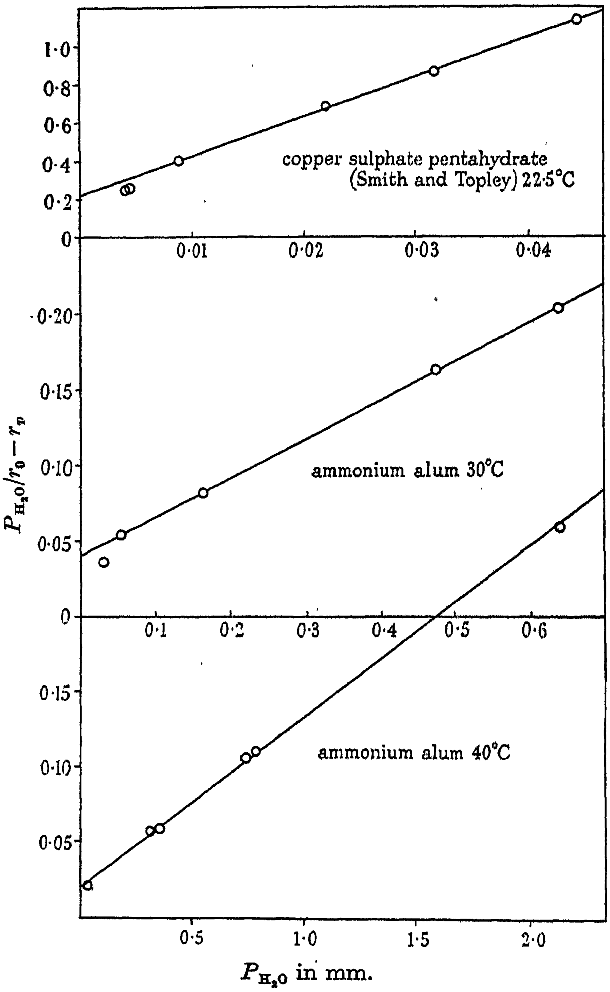


Figure 7

The rate of penetration into the crystal is definitely less than the rate of growth along the surface, but the ratio of the rates is practically constant at all temperatures and is almost identical for potassium and ammonium alums. It is unlikely that this constancy of ratio would have been obtained if there were any impedance to the escape of water vapour, since this varies with temperature. It has been shown that the rate of reaction is different in different directions in the crystal, and the slower rate of growth perpendicular to the surface is probably very largely due to this effect. Although no quantitative calculation can be made on the rate of growth into the interior of the crystal from the available data, it will be noted that on the 111 faces of ammonium and potassium alums the ratio of the minimum to the maximum rate of growth is approximately 0·7, so that ratios of the order given in tables 6 and 7 are not unexpected. The activation energy for the dehydration reactions is derived later, after correction for self-cooling of the crystal.

TABLE 6. RATE OF PENETRATION FOR AMMONIUM ALUM

temp. (°C)	self-cooling (°C)	number of molecules lost	experimental rate R_1 (cm./min. $\times 10^{-4}$)	rate from nuclei R_2 (cm./min. $\times 10^{-4}$)	R_1/R_2
20	1·1	19·8	0·35	0·71	0·50
25	1·8	19·4	0·54	0·975	0·55
30	2·75	20·0	0·76	1·415	0·53
35	4·15	20·4	1·13	2·125	0·53
40	6·20	20·6	1·42	2·68	0·53
				av.	0·53

TABLE 7. RATE OF PENETRATION FOR POTASSIUM ALUM

temp. (°C)	self-cooling	number of molecules lost	experimental rate R_1 (cm./min. $\times 10^{-4}$)	rate from nuclei R_2 (cm./min. $\times 10^{-4}$)	R_1/R_2
15	—	—	0·210	0·357	0·59
20	—	20·0	0·300	0·565	0·54
30	—	20·2	0·715	1·22	0·59
40	—	20·7	1·374	2·52	0·55
				av.	0·57

One run was carried out at 35° C for the mixed alum containing 50 % chromium and 20·87 mol. of water were lost, showing that the reaction follows a similar course to that in potassium alum: the rate of penetration was $0{\cdot}565 \times 10^{-4}$ cm./min., which was in good agreement with the value for the rate of growth of the spherical nuclei as measured on the surface, viz. $0{\cdot}556 \times 10^{-4}$ cm./min. The solid solution containing 68% chrome alum gave a loss of 15·8 to 15·9 water molecules, which indicates a change in the nature of the process of dehydration at the higher chromium content. The rate of penetration at 35° was $0{\cdot}917 \times 10^{-4}$ cm./min., which is much less than that for the radial nuclei growth. If the dehydration of the crystal was stopped half-way, it was found that the actual penetration, as observed on a broken crystal, was much greater than the value calculated from the loss of weight as measured on the spring balance. There were seen to be two layers, the inner layer purple in colour and the outer layer greenish blue. It was not possible accurately to ascertain the number of water molecules lost in the two stages, but it was deduced that approximately 8 to 12 mol. were lost in the first stage. The rate of penetration for 85·8% chrome alum could also not be accounted for except by the assumption of two successive dehydration processes. The rates of penetration as measured on the spring balance were 25° C, 1·06 cm./min. $\times 10^{-4}$ and at 30° C, 1·59 cm./min. $\times 10^{-4}$, and when measured from a broken crystal, higher values were obtained, indicating the occurrence of two processes, the first occurring with a loss of 12 water molecules and the second with a loss of 4 mol. Two stages in the process of dehydration have been previously found for copper sulphate pentahydrate (Garner & Tanner 1930).

Heats of dissociation

The dissociation pressures were measured for ammonium alum by the method previously described. Large single crystals, 2 cm. across, were employed with the interface at different depths below the surface of the crystal. The dissociation pressures were measured from 20·0 to 42° C (table 8). Log p is plotted against $1/T$ in figure 8, from which the heat of dissociation is deduced to be 15·65 cal./mol. H_2O.

TABLE 8. VAPOUR-PRESSURE MEASUREMENTS ON AMMONIUM ALUM

condition	temp. (° C)	pressure (cm.)
thick penetration	42·6	0·538
	40·1	0·484
	37·5	0·402
	35·0	0·347
	32·5	0·287
thick penetration	42·6	0·553
	40·1	0·512
	37·6	0·410
	35·1	0·350
	32·5	0·264
	30·0	0·211
	27·6	0·155
	25·5	0·128
	22·6	0·100
	42·8	0·564
	21·2	0·077
thick penetration	42·6	0·570
	30·2	0·214
	22·0	0·080
new crystals, surface nucleation only	42·4	0·550
	36·2	0·375
	20·4	0·073
	30·1	0·199
	25·5	0·136
	20·0	0·087

Self-cooling of the crystal during dehydration

The self-cooling of the crystal was measured by means of a thermocouple placed between two crystals carefully cemented together. When the reaction occurs at an interface completely surrounding the two crystals, the thermocouple will be in a constant temperature enclosure. The values for self-cooling recorded in the tables are probably somewhat too low on account of the use of too thick copper wire in the copper-constantan thermocouple, so that the self-cooling has also been calculated by the method due to Smith & Topley from the equation

$$K(T_0^4 - T_1^4) = NH,$$

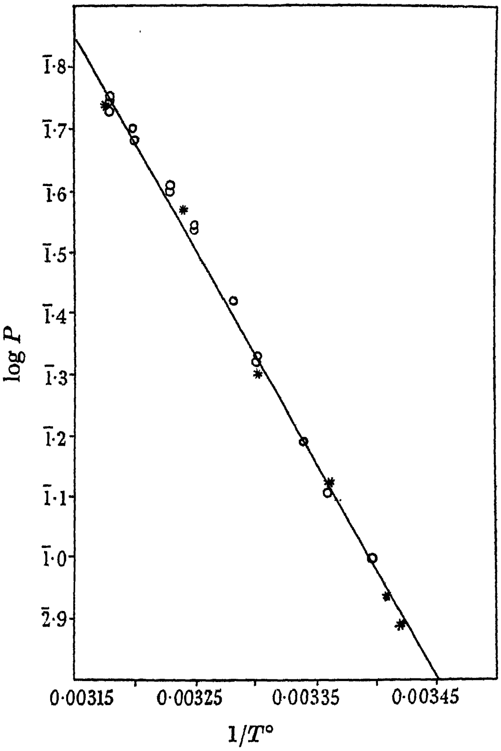


FIGURE 8. Dissociation pressures of ammonium alum.
○ deep penetration, * light penetration.

where T_0 is the temperature of the thermostat, T_1 that of the crystal, $K = 1{\cdot}37 \times 10^{-12}$, N the number of water molecules evaporated per sq.cm. and H the heat of dissociation of the hydrate. This assumes that the dehydrated product behaves as a black body, and these calculations may therefore give values for the self-cooling which are too high. Both the experimental results and those calculated assuming black body conditions were used to correct the rates of growth of nuclei and the rates of penetration, and in the calculation of the activation energies for dehydration. The values are given in table 9.

The agreement between the values for the activation energies for the nuclei growth and the penetration of the interface is satisfactory. The values for the rate of penetration would be expected to be the more reliable, since the rates were determined under conditions where the self-cooling had become constant, which is not necessarily the case for the measurements with nuclei. For ammonium alum the activation energy is about 1 kcal. higher than the energy of dissociation. The activation energy is 16·5 ± 0·5 and the heat of dissociation is 15·7 kcal. The close agreement between the two energies indicates that the coupling between the evaporation process and the reorganization of the lattice to give the final products is very close. The reformation of the alum lattice in the reverse process by the accretion of water vapour thus takes place without appreciable activation energy. The process of hydration is thus analogous to that of condensation of water vapour on a water surface, and to the reformation of carbonates from the oxides.

TABLE 9. ACTIVATION ENERGIES OF AMMONIUM AND POTASSIUM ALUMS

	activation energies	
	based on experimental self-cooling	self-cooling calculated from Planck's constant
ammonium alum		
hexagonal nuclei on 111 face	16·3	17·1
spherical nuclei on 111 face	16·0	17·1
penetration	16·4	17·2
heat of dissociation	15·7	—
potassium alum		
hexagonal nuclei on 111 face	16·2	—
penetration	16·6	—

TABLE 10

% chrome alum	activation energy (kcal.)
0·0	16·2
30·23	16·1
39·75	17·0
50·40	24·4
67·98	21–43
85·8	17–33
100·0	31

The B-factor in the equation $r_0 = Be^{-E/RT}$ for ammonium alum for the rate of reaction in molecules of water per sq.cm. of surface is $8{\cdot}2 \times 10^{12}$ at 35° and $8{\cdot}6 \times 10^{12}$ at 25° C, and is $2{\cdot}2 \times 10^{11}$ for potassium alum. These values are a little lower than those obtained in exothermic solid reactions but are in fair agreement with those expected from the Polanyi-Wigner equation, in which $B \sim 10^{13}$.

Activation energies for mixed alums. The temperature coefficients of the rates of growth for mixed alums were only determined roughly; insufficient points were taken for accurate values to be derived. The change in the character of the nuclei on the addition of chromium did not at first affect appreciably the activation energy (table 10), but when the chrome alum content exceeded 50 % the activation energy depended on the type of nuclei and on the temperature. The cluster type of nuclei had high temperature coefficients.

DISCUSSION

Nucleation. Water molecules adsorbed by van der Waal forces on the surfaces of hydrates will be removed very rapidly in vacuum, but those molecules embedded in the ionic lattice will require high energies of activation for their release. The length of the induction period before nuclei are formed and the high temperature coefficient of the induction period afford evidence of the latter. The release of this water must be accompanied by movements of the ions surrounding the water molecules, and the energy of activation is probably high because of the necessity for imparting a sufficiently large amplitude of vibration or rotation to these ions to allow the water molecule to escape. Certain of the water molecules in the alums are structural water (Lipson & Beevers 1935) and will be more mobile than the rest. The loss of a water molecule will produce a hole which will be partly filled up by some reorganization of the ions and other water molecules in its immediate neighbourhood. This hole could be completely filled by the migration of a molecule of water from within the crystal if the activation energy for migration were sufficiently low. The activation energy for migration would be expected to be of the same order as that for evaporation, so that an appreciable defect of water within the crystal lattice should be produced. There is, however, no case known where nuclei of a new solid phase are produced within the lattice. If, however, a nucleus were formed within the lattice, it would grow very slowly because there would be no rapid means for the removal of water. Also, there is more mobility for forming new solid phases at the surface than in the interior of the crystal, and therefore the absence of nuclei formation in the interior cannot be taken as evidence for the absence of lattice defects.

It is considered that the possibility of such a defect occurring within the crystal should always be borne in mind in the interpretation of the phenomena of nucleation. Certain of the phenomena found in this investigation may be interpreted in terms of a lattice defect. Thus, the length of the induction period which elapses before nuclei are formed may, in part, be due to the time needed for building up a sufficiently large water defect within the lattice. The formation of nuclei would be expected to be dependent approximately on the nth power of the concentration of the holes, where n is the number of water molecules which must be removed in one localized area in order to start the initial dislocation from which a nucleus commences to grow. Where n is large, nuclei will form when the concentration of defects approaches a critical value and the time at which nucleation commences will be reasonably sharp, as is, in fact, found. From the counts of the numbers of nuclei of different sizes present at any time, it was deduced that the rate at which the nuclei were formed varies according to a linear or a second order law (Bright & Garner 1934; Garner & Southon 1935; Cooper & Garner 1936). This is what would be expected if, during the period in which nuclei were forming, the water defect within the crystal was being maintained at a constant or slightly increasing figure. The defect within the lattice would be increased by the loss of water from the crystal surface and decreased by migration of the holes to the interface of the nuclei. When the area of the interface becomes large, evaporation from the bare crystal surface may be insufficient to maintain the lattice defect above the critical value at which nuclei are readily formed. Therefore, after an initial batch of nuclei has been formed, it would be expected that there would be a complete cessation of further nucleation. This is actually observed in the cases of common and ammonium alums. On the basis of this reconstruction, the initial rapid rate of growth of the nuclei (figure 5) can be accounted for as due to the migration of lattice defects to the growing interface, which will increase temporarily the rate of loss of water from this interface.

Interface reaction and shapes of nuclei. The striking fact about interface reactions is that the removal of water is much easier from the interface than from the crystal surface. This is probably due to the coupling of the process of evaporation with an exothermic process, the reorganization of the inorganic constituents to give new solid phases. There appear to be two types of interface reaction, possibly linked by intermediate cases. In the first class, of which copper sulphate pentahydrate, potassium and ammonium alum are examples, the activation energies are those normal for processes occurring at room temperatures and the B-factor in the equation $r_0 = Be^{-E/RT}$ is of the order of 10^{11} to 10^{13}. The heat of activation is of the same order as the thermodynamic heat of dissociation. In the second class, of which chrome alum and calcium carbonate hexahydrate are examples, the activation energy is considerably higher than the heat of dissociation and the B-factor many powers of 10 higher than the normal value. The nature of the coupling between the evaporation of water and the reorganization of the lattice may be different in the two cases.

Normal interface reactions. For these the rate of removal of water from the interface is markedly dependent on crystal direction, and the nuclei possess characteristic geometrical shapes which reproduce some of the symmetry of the crystal. The processes which occur may be regarded as a reversal of those occurring during the growth of a crystal from vapour or solution, in which the crystal is built up along selected planes (Volmer 1939). By analogy with the growth and dissolution of crystals, the activation energy for the removal of water from any point on the interface would be expected to depend on the number of undecomposed units surrounding that point. Thus the loss of water from the middle of a plane surface will be less probable than that of a molecule placed at a corner or along an edge. The interface is not, however, completely analogous to a surface, for there is the effect of the second solid phase to be taken into account, which might modify the activation energies. The shape of the nucleus will, however, depend on the relative values of the activation energies of the various types of process that are permissible. If the activation energies are all approximately of the same order, then no crystal direction will be specially preferred and a spherical nucleus will be obtained. In the case of potassium and ammonium alums the bounding of the surfaces of the nuclei by edges which lie in the 100 planes of the hydrate indicates that the process of dehydration is fastest along the 100 planes and that the process occurs layer by layer. It would appear that phenomena observed in the dehydration of these two alums are normal from a kinetic point of view because of this fact, since layer decomposition would facilitate a close coupling between the dehydration process and the reorganization of the ions to give a new lattice.

For the normal mechanism of coupling, the rate is retarded by water vapour, and the probable mechanism of the retardation is that the adsorption of water molecules on active areas in the interface blocks the dehydration. The rate of dehydration becomes proportional to the number of such places not covered by adsorbed water. In the presence of water vapour, the nuclei lose their characteristic shapes and become approximately spherical, indicating that layer growth is hindered, and that the reaction takes other paths which possess equal probability in all directions. When the water vapour pressure is just below the dissociation pressure, the nuclei possess an irregular outline and become coarsely crystalline. They are similar to the nuclei obtained with iron ammonium alum in a hard vacuum and resemble the spherulitic growths sometimes formed from solution, where there is a very rapid growth of the crystals along one axis. Layer growth has obviously broken down into a process which is determined more by the properties of the crystalline products than by the structure of the original hydrate.

The addition of foreign ions to the hydrate lattice also causes a breakdown of the layer growth, and produces spherical nuclei. The effects are, however, radically different from those due to the presence of water vapour. The foreign ions appear to stabilize the lattice, conferring on it some of the properties of a zeolitic structure. There is evidence that water is removed from the region ahead of the normal interface, which indicates that the movement of holes within the lattice is facilitated. The occurrence of such a process would obviously cause a breakdown of dehydration layer by layer.

There is a diffuse region ahead of the interface in the case of ammonium and potassium alums (figure 1 *a*, *b*, plate 12), which appears to consist of fine cracks. This phenomenon only leads to second order effects in the case of the two alums, and the layer growth controls the phenomena which occur. The formation of cracks ahead of the interface might, under certain conditions, have serious effects, for if the cracks isolated portions of the crystals, these would have to be nucleated afresh, and the rate of dehydration would be controlled by nucleation in the cracked area.

Abnormal interface reactions. These are characterized by high values of E and high B-factors. A high activation energy would be expected if the process of evaporation were not closely coupled with the formation of crystals of the solid products.

Thus it would be high if the products were produced in an amorphous form. Chrome alum gives $E = 31$ kcal. and calcium carbonate hexahydrate $E = 38$ to 40 kcal., and these values are so high that it would not be expected that the reactions would occur at an appreciable rate at room temperatures. For chrome alum $r_0 = 10^{25} . e^{-31,000/RT}$ mol./cm.^{2}, and this rate could be accounted for if the excitation of a molecule with 31 kcal. sufficed for the release of 10^{12} other molecules. This would correspond with the nucleation of a block of a crystal 10^{-3} cm. cube. It may well be that for chrome alum, cracking occurs ahead of the interface, producing isolated blocks of crystal which have to be nucleated afresh. If this were the rate-determining step, the high activation energy would be accounted for.

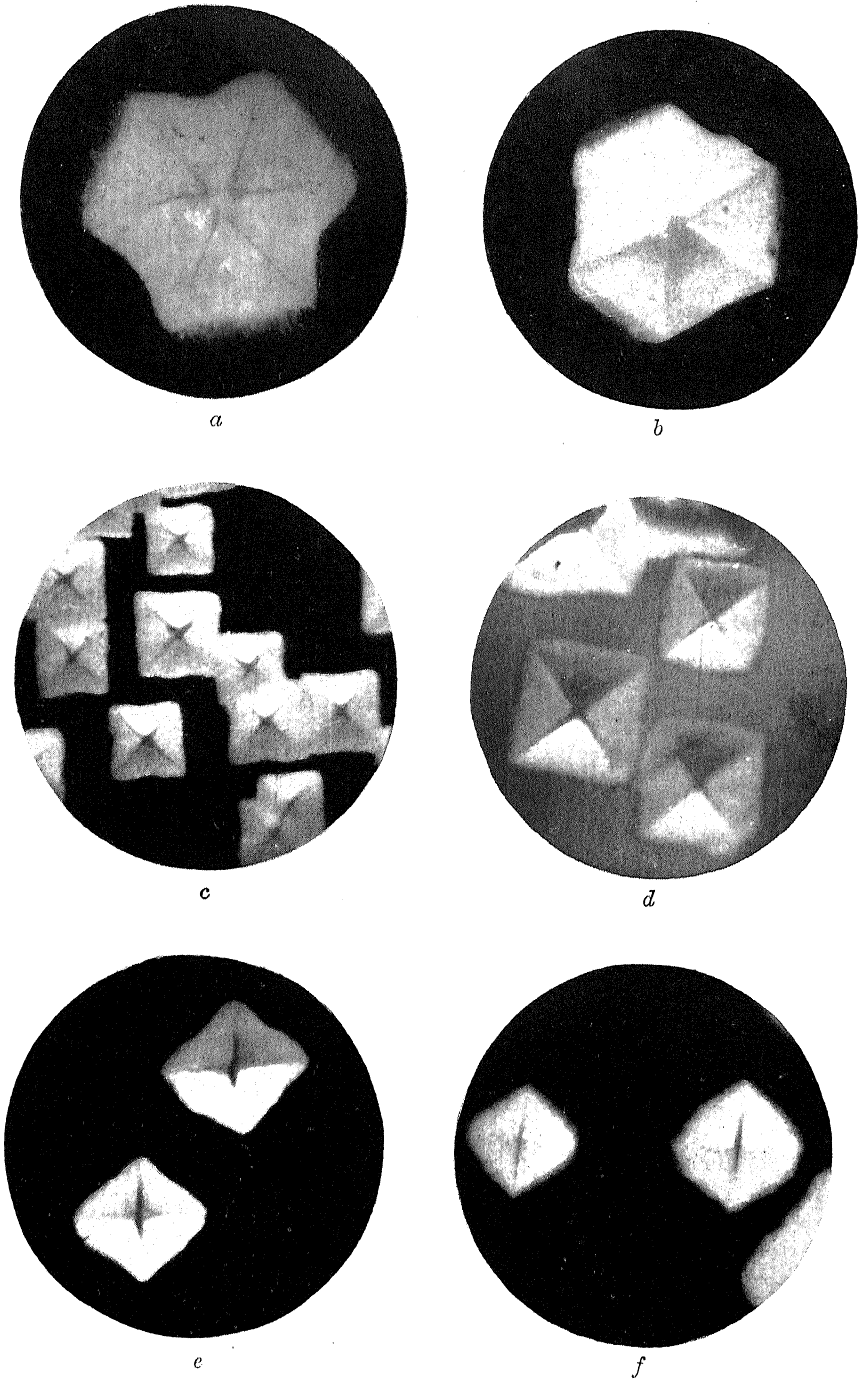
It has been suggested previously (Garner 1939) that the activation was a co-operative process, proceeding at a number of points in the lattice simultaneously, thereby producing a dislocation over a wider area than with the normal coupling. This would mean breaking down the activation energy into a number of smaller values, applied at a number of points in the interface. This might be the correct explanation.

There is a third possibility, that an amorphous product is formed in a transition layer which hinders the flow of water molecules from the interface, and that the thickness of this layer is dependent on temperature. This would mean that the high activation energy was fictitious, and that there should be subtracted from it the energy corresponding to the temperature coefficient of the impedance through the transition layer. This would also give a reasonable explanation.

It does not appear to be possible at the present stage to decide which of these explanations are correct.

REFERENCES

Bright & Garner 1934 *J. Chem. Soc.* p. 1872.
Cooper & Garner 1936 *Trans. Faraday Soc.* **32**, 1739.
Cooper & Garner 1940 *Proc. Roy. Soc.* A, **174**, 487.
Garner 1939 *Nature*, **144**, 287.
Garner & Pike 1937 *J. Chem. Soc.* p. 1565.
Garner & Southon 1935 *J. Chem. Soc.* p. 1705.
Garner & Tanner 1930 *J. Chem. Soc.* p. 47.
Hume & Colvin 1929 *Proc. Roy. Soc.* A, **125**, 635.
Kohlschütter & Nitschmann 1931 *Z. phys. Chem.*, Bodenstein Festband, p. 494.
Lipson & Beevers 1935 *Proc. Roy. Soc.* A, **148**, 680.
Smith & Topley 1931 *Proc. Roy. Soc.* A, **134**, 224.
Smith & Topley 1935 *J. Chem. Soc.* p. 321.
Spencer & Topley 1929 *J. Chem. Soc.* p. 2645.
Topley 1932*a* *Proc. Roy. Soc.* A, **136**, 413.
Topley 1932*b* *Phil. Mag.* (VII), **14**, 1080.
Topley & Hume 1928 *Proc. Roy. Soc.* A, **120**, 211.
Volmer & Seidel 1937 *Z. phys. Chem.* **179**, 153.
Volmer 1939 *Kinetik der Phasenbildung*, Steinkopf.
Zawadzki & Bretsznajder 1938 *Trans. Faraday Soc.* **34**, 951.

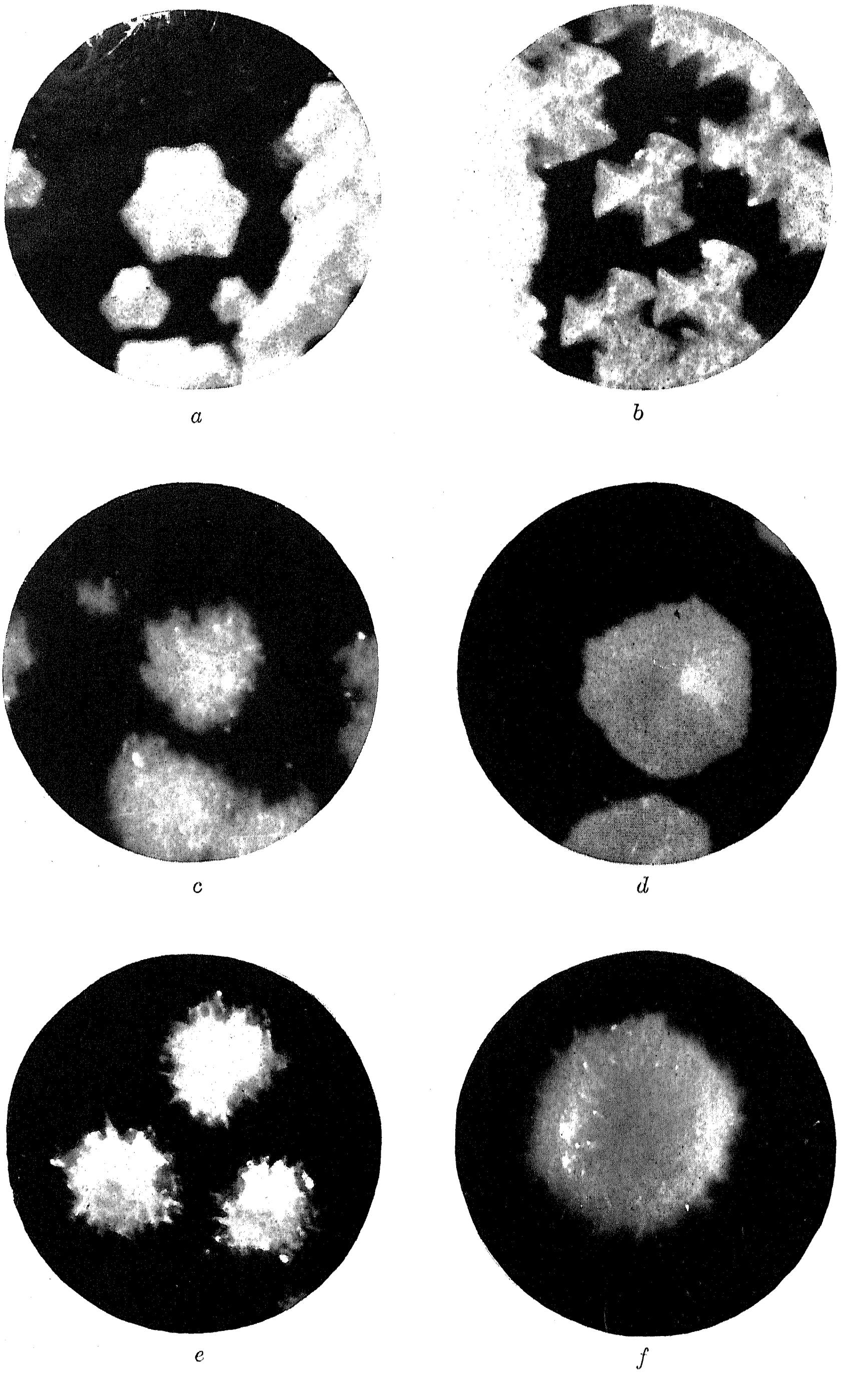


a *b*

c *d*

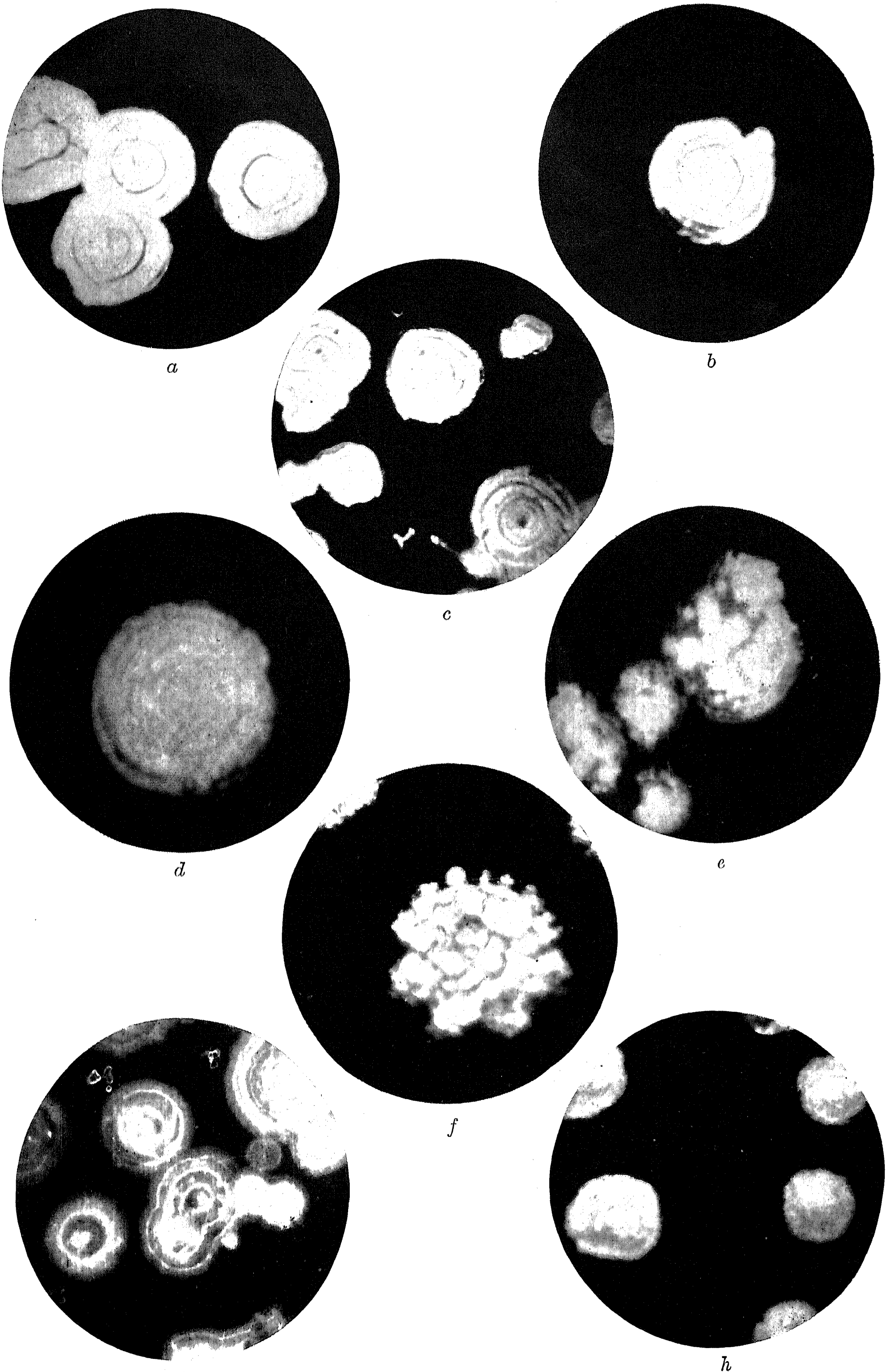
e *f*

FIGURE 1



a *b* *c* *d* *e* *f*

FIGURE 3



a *b* *c* *d* *e* *f* *g* *h*

FIGURE 4

DESCRIPTION OF PLATES 12 to 14

PLATE 12. FIGURE 1. Nuclei on potassium and ammonium alums.

potassium alum		ammonium alum	
a	111 face	*b*	111 face
c	001 face	*d*	001 face
e	011 face	*f*	011 face

PLATE 13. FIGURE 3. Nuclei on alums.

a Rubidium alum, 111 face.
b Caesium alum, 111 face.
c Iron ammonium alum, 111 face.
d Ammonium alum, 111 face. Effect of 0·776 mm. water vapour during growth at 40° C.
e Ammonium alum, 111 face. Effect of 2·05 mm. water vapour during growth at 40° C.
f Potassium alum, 111 face. Effect of 0·776 mm. water vapour during growth at 30° C.

PLATE 14. FIGURE 4. Nuclei on $K_2SO_4[Al, Cr](SO_4)_3.24H_2O$.

Slow neutron-induced activity in gold

BY K. SIEGBAHN, *Research Institute for Physics, Academy of Sciences, Stockholm*

(*Communicated by Sir Owen Richardson, F.R.S.—Received* 20 *March* 1946)

The 2·7 day activity in ^{198}Au has been studied by means of a β-spectrometer of the lens type in conjunction with a special technique for studying the γ-radiation. The β-spectrum has an upper limit = 0·92 MeV. An absorption measurement on the γ-radiation in Pb shows that no high-energy γ-radiation is existent, as has previously been reported. Only one γ-ray of medium energy (0·401 MeV) follows the β-disintegration.

If gold is irradiated with slow neutrons (Fermi, Amaldi, d'Agostino, Rasetti & Segrè 1934), a strong activity appears with a half-life period of 2·7 days. This must be due to the gold isotope 198, as gold has only one stable isotope with the mass 197. The same active isotope can be produced in several other different ways, for example, by deuteron-irradiation of gold (Cork & Thornton 1937) and platinum (Lawson & Cork 1940). At the same time, however, other active gold isotopes are produced, and hence the methods mentioned are not suitable if it is desired to study ^{198}Au alone. Fast neutrons, on the other hand, give rise, besides the above-mentioned activity, to two others, though considerably weaker, with the half-life periods 13 hr. and 3·3 days (MacMillan, Kamen & Ruben 1937).

Several investigations of the disintegration of ^{198}Au have already been performed, different methods being used: absorption measurements, coincidence measurements, investigations in the Wilson chamber and the β-spectrograph. The results have hitherto shown little mutual agreement, which, in some cases, may be explained by the presence of active isotopes other than ^{198}Au. A survey of earlier measurements has been published by Feather & Dainty (1944). For the sake of completeness some of the essential results will be repeated here.

^{198}Au emits both β- and γ-radiation. It is transformed by β-emission into ^{198}Hg. Sizoo & Ejkman (1936) have, however, found, in absorption measurements on the γ-radiation, a weak γ-energy at 2·5 MeV. This is explained by assuming that ^{198}Au may also be transformed into ^{198}Pt by K-capture.

The β-spectrum has been studied by, among others, Richardson (1939) in a Wilson chamber. The sample consisted of a gold foil that had been irradiated with slow neutrons. The spectrum proved to consist of a highly asymmetrical, continuous distribution, with its upper limit at 0·83 MeV. Besides, a strong β-line was found, corresponding to an internally converted γ-energy of 440 keV. The conversion coefficient was as high as 0·1. The upper limit of the β-spectrum has been further studied by Sizoo & Ejkman, using the absorption method. The result $E_{\text{max.}} = 0{\cdot}90$ MeV was obtained. Using the same method, Cork & Halpern (1940), among others, obtained the result $E_{\text{max.}} = 0{\cdot}74$ MeV and Clark (1942) $E_{\text{max}} = 0{\cdot}78$ MeV (corrected by Feather & Dainty (1944) to 0·87 MeV). Using a particularly exact absorption method (Feather 1938), Feather & Dainty determined $E_{\text{max.}}$ $0{\cdot}985 \pm 0{\cdot}010$ MeV.

The γ-radiation accompanying the disintegration is of special interest. In several different cases, two γ-lines were discovered, with energies of about 250 and 450 keV respectively. At the same time, a weak radiation of about 70 keV was obtained. The latter radiation has been supposed to be X-radiation from ^{198}Hg (K-radiation). This explanation may be considered plausible, in view of the existence of the strong β-line. The following investigation on the γ-radiation may be given brief mention. Richardson (1939) determined, in a Wilson chamber, the distribution of the photoelectrons ejected by the γ-radiation from a lead foil. He obtained three photo-lines, corresponding to the γ-energies 70, 280 and 440 keV. A slight trace of yet another γ-energy of about 0·7 MeV was also obtained. The ratio between the intensities for γ_{280} and γ_{440} was estimated as 1·0 : 1·2. Using the absorption method, Sizoo & Ejkman (1936) obtained the γ-energies 73, 250 and 410 keV. Besides, they obtained a weak γ-radiation of about 2·5 MeV (about 2 quanta per 100 disintegrational). The two medium soft γ-components were also found by Cork & Halpern (1940) (240 and 500 keV) and Krishnan (1941) (210 and 500 keV). Plesset (1942) studied the γ-radiation from gold, irradiated with deuterons of 11 MeV, lead being the secondary radiator. He obtained a number of photo-lines, three of which were attributed to ^{198}Au. The photo-lines had the energies 63·9, 323 and 389 keV. As the difference between the two latter corresponds to the difference between the K- and L-absorptions energies of lead, the origin of both lines may be one γ-line with the energy 402 keV. On the other hand, the intensities of the two lines were nearly equal, which, our supposition being right, would not be the case. (The ratio should be about 7 : 1.)

A coincidence investigation by Clark (1942) showed that γ-γ as well as β-γ coincidences were obtained. By the use of Norling's (1941) efficiency curve for lead γ-γ-tubes, the coincidence measurements could be accounted for on the basis of Richardson's disintegration scheme (figure 1) which assumes that the two γ-rays are emitted in cascade and that each β-particle is accompanied by two γ-quanta (disregarding the suggested possibility of an extremely weak γ-transition direct to the ground state). Clark's coincidence measurements seem quantitatively to agree very well with the corresponding investigations of Feather & Dainty (1944). Norling (1941), on the other hand, could not obtain any γ-γ coincidence effect. On the other hand, he obtained a very strong β-γ coincidence effect, which undoubtedly indicates that at low energies, the γ-efficiency curve passes through a maximum (see p. 55 in Norling's paper).

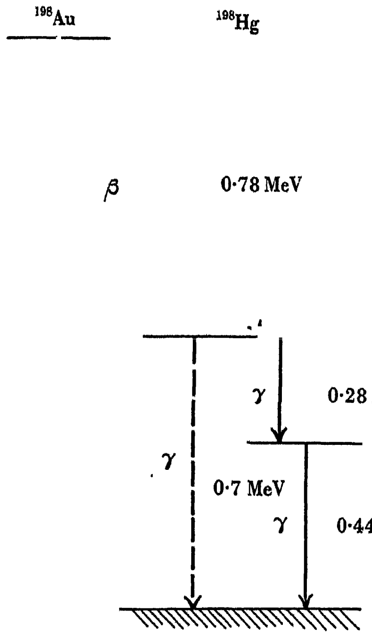


FIGURE 1. Previously proposed level scheme for the disintegration of ^{198}Au.

THE β-SPECTRUM

Since at present there is no method, admitting of a substantial concentration of active in relation to inactive gold, gold in the form of a thin foil (0·02 mm. thick) was irradiated with slow neutrons from the Stockholm cyclotron. A sample of such a thickness cannot, of course, be expected to give a β-spectrum other than with considerable distortion. The foil, whose diameter was 8 mm., was placed in the

β-lens spectrograph (Siegbahn 1946), the resolving power of which was at a value of $\Delta H\rho/H\rho = 5\%$. At this adjustment of the spectrograph, the diaphragm system transmits 2 % of the radiation. With the spectrograph so adjusted it has proved possible, with preparations of sufficiently long half-life period, to measure the β-spectrum at intensities as small as some hundredths μC.

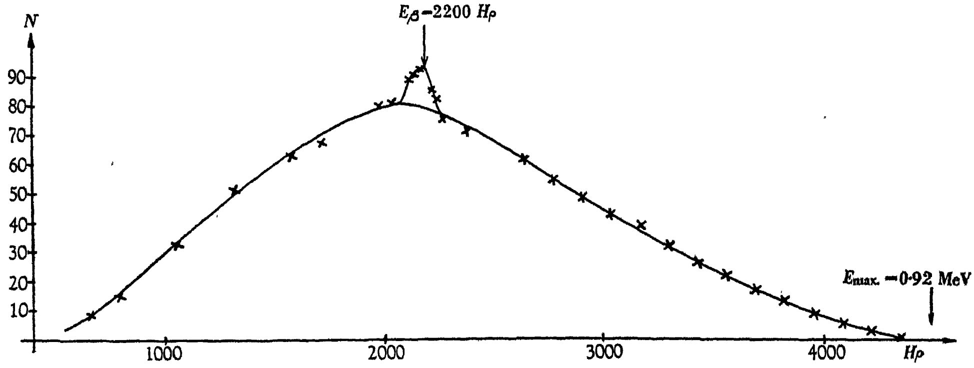


FIGURE 2. The β-spectrum of ^{198}Au.

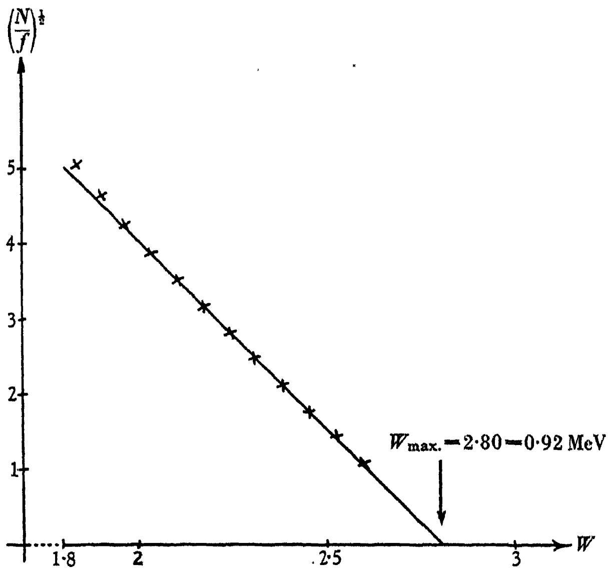


FIGURE 3. The Fermi diagram for ^{198}Au.

Figure 2 shows the β-spectrum of ^{198}Au. It is simple in character, as seen from the relatively symmetrical form. At low energies, however, it can be observed that the absorption in the gold foil has become perceptible, the spectrum displaying a comparatively small number of low-energy electrons. The spectrum shows only one β-line, with $H\rho = 2200$, corresponding to an energy for the γ-quantum = 0·401 MeV. The corresponding L-line $H\rho$ 2470 was also observable but has been omitted, as the statistics employed hardly permitted any definite conclusion regarding its magnitude. The size of the β_K-line relative to the remaining continuous spectrum is smaller than that previously observed, which may in some part be due to the absorption in the gold foil. According to Richardson's determination, the β-line should correspond to an internal conversion coefficient α_K of 10 %, whereas Plesset finds $\alpha_K = 4$ %. The uncorrected value of α_K is found from figure 2 to be about 1 %.

The upper limit of the spectrum is most accurately obtained by finding the Fermi curve for the spectrum. Figure 3 shows the end of this curve. The upper limit is determined from this as $0·92 \pm 0·01$ MeV.

Even if the β-spectrum reveals only one γ-line, there are possibilities of more than one such line occurring through disintegration. The high value of α_K for the β-line at 0·401 MeV is probably due to this radiation having a quadrupole character. If there are other γ-lines of a dipole nature, these are not certain to bring about internal conversion to an extent large enough to give detectable β-lines. It is therefore necessary to study the γ-radiation directly.

γ-ABSORPTION IN LEAD

In the first place, the question arose of ascertaining, by an absorption measurement, the presence of the weak, hard γ-component of 2·5 MeV, reported by Sizoo & Ejkman (1936). For this purpose a γ-tube was used, the cathode being a 1 mm. thick brass cylinder. The absorption process was observed for absorption thicknesses up to 28 mm. of lead. Figure 4 shows, in a logarithmic scale, the curve obtained. The intensity, at the last point measured, is only about 4 % of the intensity without absorption screens. Nevertheless, even these last points lie entirely within the limits of error on the straight line that may be drawn through the other points, and hence the determination does not support the presence of any hard component. This is in agreement with the conclusion of Feather & Dainty (1944) based on less extensive evidence. Since it was necessary to extend the absorption measurements to such large absorption thicknesses in order to establish the presence of the hard component, the sample could not be placed, owing to difficulties of intensity, as far from the GM tube as would be desirable for a good quantitative determination of the energy of the γ-radiation. The crude γ-ray absorption coefficient corresponds to an energy = 0·46 MeV. This value is considerably higher than the more accurately determined $E_\gamma = 0·401$ MeV from the β-spectrum.

One cannot expect to establish from the absorption curve, considering the thickness of the wall of the γ-tube, presence of the γ-ray of 70 keV reported by several authors. It is remarkable, however, that the γ-component of about 250 keV, also often reported, cannot be detected by any procedure. With regard to the geometry not being ideal, it is, however, difficult to draw any definite conclusion from the absorption curve as to the non-existence of this γ-energy. It is certainly known that it is sometimes possible to deduce a purely exponential absorption process, even on occasions when several γ-lines occur (Roberts, Downing & Deutsch 1941).

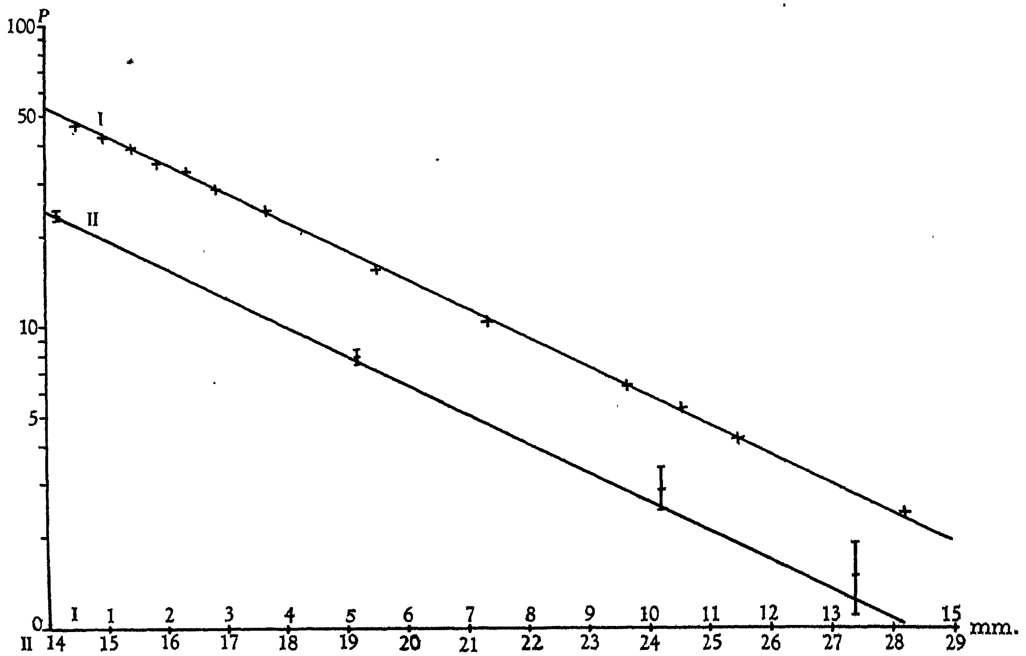


Figure 4. Absorption curve for the γ-radiation of ^{198}Au in lead.

THE STUDY OF THE γ-RAYS IN THE SPECTROGRAPH

In order to investigate this question closer, it is suitable to apply the more accurate photo-Compton method with the aid of the β-spectrograph. Gold in the form of 0·02 mm. foils was irradiated in the cyclotron with slow neutrons for a few days, so that the total activity was equivalent to $10\,\mu$C Ra. The gold was then packed into a small cylindrical Cu radiator ($\phi = 8$ mm.), the wall thickness of which was sufficiently large, according to Feather's formula, to absorb entirely the electrons of the continuous β-spectrum. The energy distribution of the secondary electrons released in the Cu radiator was studied in the β-spectrograph. A 0·1 mm. thin lead foil was then placed in front of the Cu radiator, and a new energy distribution curve was recorded. In the first case, it is fairly certain that the Compton effect in the copper provides the main component in the secondary spectrum and in the latter case both the Compton effect and the photo-effect in the lead foil will be seen.

Figure 5 shows the former registration. In order to give an idea of the sensitivity of the arrangement, I have not plotted along the axis, in this figure and figure 6, the intensity, i.e. the number of impulses counted per min. P divided by the corresponding $H\rho$-value, but P alone. Figure 5 suggests at first sight the existence of two γ-components, the lower energy corresponding to that first obtained from the

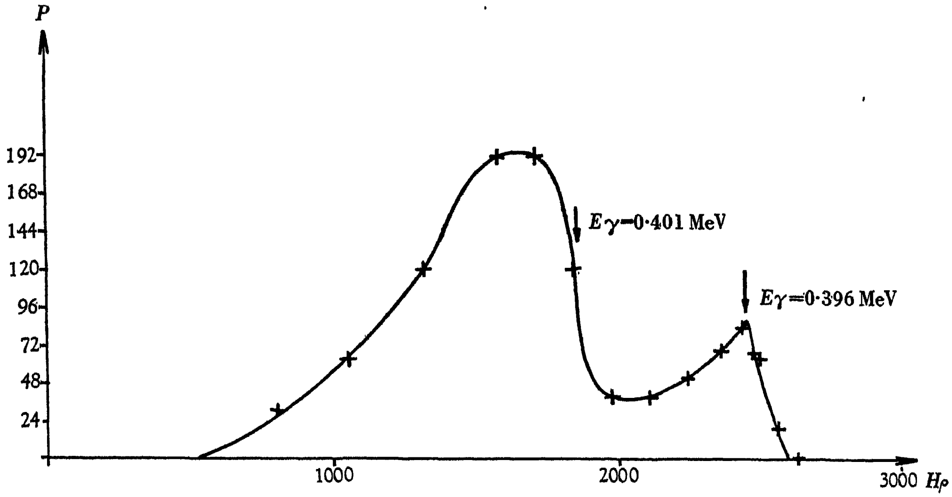


FIGURE 5. Secondary electron spectrum expelled from a Cu radiator by the γ-radiation of ^{198}Au.

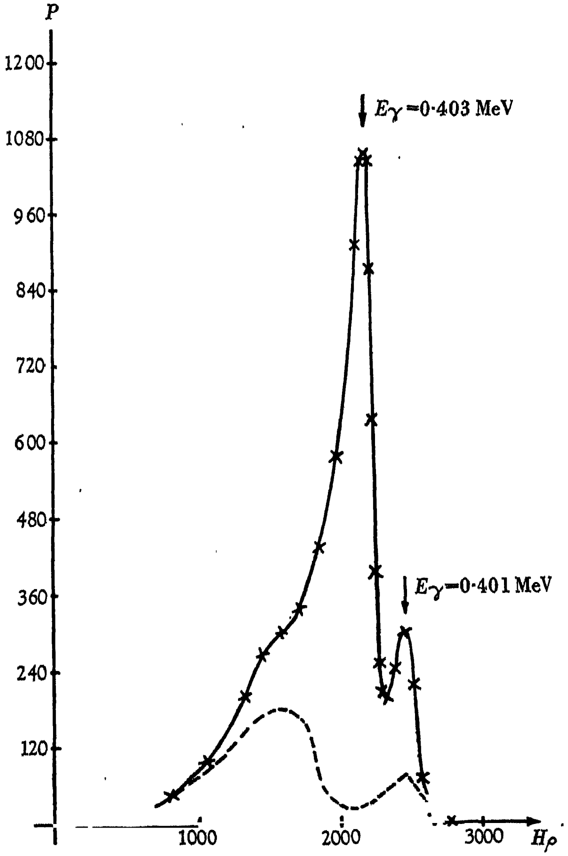


FIGURE 6. Secondary electron spectrum expelled from a Cu radiator covered with a 0·1 mm. lead foil.

β-spectrum. The next figure, however, makes it clear that the second energy peak is due to the photo-effect in the copper, which, though certainly weak, is quite perceptible at these comparatively low energies of the γ-radiation. The present author has previously found, in several cases, that the most accurate possible value is obtained for the γ-energy, if the upper energy limit of the Compton distribution (i.e. the limit corresponding to the impact angle $\theta = 0$ in the Compton effect formula) is placed at the inflexion point of the energy distribution. Using this procedure, the value $E_\gamma = 0{\cdot}401$ MeV is obtained from figure 5. The energy of the γ-radiation may also be determined from the peak of the photo-line. This has the $H\rho$-value 2484, i.e. 0·387 MeV. If the work of liberation for the K shell in Cu, 9 keV, is added, then $E_\gamma = 0{\cdot}396$ MeV.

Figure 6 shows the distribution of the secondary electrons, when the lead foil was placed before the radiator. This curve is composed of three different components: the Compton electrons (the broken line represents the curve of figure 5 for comparison), the K photo-line in Pb and the L photo-line in Pb. It is evident that the higher energy maximum in figure 5 cannot be a Compton distribution from a second γ-component, as the photo-line that would correspond to this γ-energy in figure 6 is missing. The peak of the K photo-line occurs at $H\rho = 2161 = 0{\cdot}315$ MeV. Hence E_γ is obtained by adding the K absorption energy of the lead $= 0{\cdot}088$ MeV, i.e. $E_\gamma = 0{\cdot}403$ MeV. E is obtained in the same way from the L photo-line, the $H\rho$-value of which is 2451 as ($E_{L_{\text{abs.}}} = 0{\cdot}016$ MeV) 0·401 MeV. It is satisfactory that the energy values deduced from the internal and external conversion effects agree so closely, and it should further be noted that the final value of the γ-ray energy deduced from the present experiments are in excellent accord with the value obtained by Plesset (1942). It is further satisfactory that the strange relation between the intensities of the two photo-lines reported by Plesset does not appear in the present investigation.

Figure 6 illustrates well the sensity of the spectrograph. It proves, among other things, that the activity of only 1 μC Ra γ equivalence would still have given about 100 kicks/min. at the peak of the K photo-line. Since the zero effect of the GM tube gives about 5 kicks/min., it appears possible to use, within this energy range, still weaker samples for recording γ-spectra.

The conclusion drawn from figures 5 and 6 is that there is only one γ-line, and, accordingly, that the γ-line at about 250 keV, earlier reported by several authors, must be regarded as non-existent. The present investigation is not concerned with the very low energy range at about 70 keV, in which a γ-line has earlier been reported. As pointed out earlier this was considered to be X-radiation from mercury, being the result of internal conversion from the γ-energy at 401 keV. Feather & Dainty (1944), however, have made a closer study of the absorption process in this energy range and seem to have succeeded in proving the presence of a nuclear γ-line, the energy of which is somewhat lower than that of the X-ray, i.e. about 65 keV. If the existence of a nuclear γ-radiation of 65 keV is accepted, there are two possibilities for its emission. It may be emitted after the whole β-spectrum or it may be responsible for a complexity in the β-spectrum, the latter being composed by two components differing by 65 keV in energy. In the first case the two γ-lines must be of equal intensity. Very carefully performed coincidence experiments might probably give some further information on this point.

REFERENCES

Clark, A. 1942 *Phys. Rev.* **61**, 242.
Cork, J. M. & Halpern, J. 1940 *Phys. Rev.* **58**, 201 A.
Cork, J. M. & Thornton, R. L. 1937 *Phys. Rev.* **51**, 59.
Feather, N. 1938 *Proc. Camb. Phil. Soc.* **34**, 599.
Feather, N. & Dainty, J. 1944 *Proc. Camb. Phil. Soc.* **40**, 57.
Fermi, E., Amaldi, E., d'Agostino, O., Rasetti, F. & Segrè, E. 1934 *Proc. Roy. Soc.* A, **146**, 483.
Krishnan, R. S. 1941 *Proc. Camb. Phil. Soc.* **37**, 186.
Lawson, J. L. & Cork, J. M. 1940 *Phys. Rev.* **58**, 580.
MacMillan, E., Kamen, M. & Ruben, S. 1937 *Phys. Rev.* **52**, 375.
Norling, F. 1941 *Ark. Mat. Astr. Fys.* A, **27**, no. 27.
Plesset, E. H. 1942 *Phys. Rev.* **62**, 181.
Richardson, J. 1939 *Phys. Rev.* **55**, 609.
Roberts, A., Downing, J. R. & Deutsch, M. 1941 *Phys. Rev.* **60**, 544.
Siegbahn, K. 1946 *Phil. Mag.* **37**, 162.
Sizoo, G. J. & Ejkman, G. 1936 *Physica, Eindhoven*, **6**, 332.

Mixed boundary conditions in the relaxational treatment of biharmonic problems (plane strain or stress)

BY L. FOX, D.PHIL.

(*Communicated by R. V. Southwell, F.R.S.—Received* 6 *July* 1946)

Relaxation methods have already been applied to the solution of four problems of (i) extension and (ii) flexure of flat elastic plates, in which (*a*) displacement or (*b*) traction is specified at the boundary. Here the method is adapted to the case in which the two types of boundary condition are mixed, where photo-elastic methods are difficult to apply.

Two examples are treated by relaxation methods, and the results obtained indicate that this method may be a valuable alternative in engineering problems.

INTRODUCTION

1. Part VII A of the series 'Relaxation methods applied to engineering problems' (Fox & Southwell 1941), was concerned with the solution of problems involving flexure or extension of flat elastic plates, for which the governing equations are

biharmonic. Four cases were considered, differing in respect of the boundary conditions, namely:

1 (*a*). Problems of extension in which the boundary conditions specify displacements u, v in two perpendicular directions.

1 (*b*) Problems of flexure in which the boundary conditions specify the boundary tractions.

2 (*a*). Problems of extension in which the boundary tractions are specified.

2 (*b*). Problems of flexure in which the boundary displacements are specified.

The governing equations for Case 1 (*a*) are (in plane stress)

$$\left.\begin{aligned}\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1-\sigma}{1+\sigma}\left(\nabla^2 u+\frac{\rho X}{\mu}\right)=0,\\ \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1-\sigma}{1+\sigma}\left(\nabla^2 v+\frac{\rho Y}{\mu}\right)=0,\end{aligned}\right\} \tag{1}$$

where X, Y are body forces. In Case 2 (*b*) a single equation

$$\nabla^4 w = Z \tag{2}$$

relates the transverse displacement w with the intensity Z of applied transverse loading. It was shown in the paper cited that problem 1 (*b*) can be formulated in equations of the type (1), and that a similar analogy holds between problems 2 (*a*), 2 (*b*).

2. When the boundary conditions are 'mixed' (displacements specified at some points, tractions at others), the problem is much more difficult both for analysis and for (photo-elastic) experiment. A technique for solving problems of this kind would accordingly have value, and an accuracy of, say, 5 % would be sufficient for most engineering purposes. The boundary conditions, where they specify tractions, present difficulties, especially at a curved boundary; but recent developments of the relaxation method have gone some way towards overcoming these difficulties, though at some cost in accuracy compared with what is attainable when the boundaries are rectangular.

Here, two problems of the kind are solved: the first, an example treated by Timoshenko (1934, § 43) in which the boundary is a square; the second, an example having a circular boundary, of which the analytical solution is known. Satisfactory results are obtained in both.

DESCRIPTION OF METHOD

3. In the absence of body forces the governing equations are

$$\left.\begin{aligned}\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1-\sigma}{1+\sigma}\nabla^2 u=0,\\ \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1-\sigma}{1+\sigma}\nabla^2 v=0.\end{aligned}\right\} \tag{3}$$

On the boundary, u and v are given at some points; the boundary traction at others, i.e. specified values of

$$X_\nu = X_x \cos(x,\nu) + X_y \sin(x,\nu),$$
$$Y_\nu = X_y \cos(x,\nu) + Y_y \sin(x,\nu).$$

Expressing these in terms of u, v by means of the stress-strain relations

$$\left.\begin{aligned} kX_x &= \frac{\partial u}{\partial x} + \sigma\frac{\partial v}{\partial y}, \\ kY_y &= \frac{\partial v}{\partial y} + \sigma\frac{\partial u}{\partial x}, \\ kX_y &= \frac{1-\sigma}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right), \end{aligned}\right\} \quad \left(k = \frac{1-\sigma^2}{E}\right) \tag{4}$$

we obtain

$$\left.\begin{aligned} 2\frac{(1-\sigma^2)}{E}X_\nu &= 2\left(\frac{\partial u}{\partial x} + \sigma\frac{\partial v}{\partial y}\right)\cos(x,\nu) + (1-\sigma)\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\sin(x,\nu), \\ 2\frac{(1-\sigma^2)}{E}Y_\nu &= 2\left(\frac{\partial v}{\partial y} + \sigma\frac{\partial u}{\partial x}\right)\sin(x,\nu) + (1-\sigma)\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\cos(x,\nu), \end{aligned}\right\} \tag{5}$$

as boundary conditions for points where X_ν, Y_ν are specified.

Together with the governing equations (3), the boundary conditions (thus expressed in u and v) define the problem completely. They could be used for problems of class 2 (a), in which boundary traction is specified everywhere; but the method of Part VII A is usually simpler and accordingly more accurate.

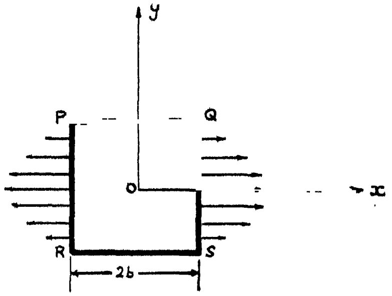


Figure 1

As an example we take the problem considered by Timoshenko: a square plate of side $2b$ is stretched by forces acting on two opposite ends, the forces being distributed according to a parabolic law (figure 1). Here the boundary conditions (5) are simple: we have

$$(x,\nu) = \tfrac{1}{2}\pi \text{ on } PQ,\ 0 \text{ on } QS,$$
$$X_\nu = (1 - y^2/b^2) \text{ on } QS,\ 0 \text{ on } PQ,$$
$$Y_\nu = 0 \text{ on all boundaries,}$$

and thus equations (5) reduce to

$$\left.\begin{aligned}\frac{1-\sigma^2}{E}\left(1-\frac{y^2}{b^2}\right) &= \frac{\partial u}{\partial x}+\sigma\frac{\partial v}{\partial y}\\ 0 &= \frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\end{aligned}\right\}\quad \text{on } QS, \tag{6}$$

$$\left.\begin{aligned}0 &= \frac{\partial v}{\partial y}+\sigma\frac{\partial u}{\partial x}\\ 0 &= \frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\end{aligned}\right\}\quad \text{on } PQ. \tag{7}$$

The finite-difference equivalents of (3) at the point O in figure 2 are

$$\left.\begin{aligned}2(u_1+u_3-2u_0)+\frac{1+\sigma}{4}(v_5-v_6+v_7-v_8)+(1-\sigma)(u_2+u_4-2u_0) &= 0,\\ 2(v_2+v_4-2v_0)+\frac{1+\sigma}{4}(u_5-u_6+u_7-u_8)+(1-\sigma)(v_1+v_3-2v_0) &= 0.\end{aligned}\right\} \tag{8}$$

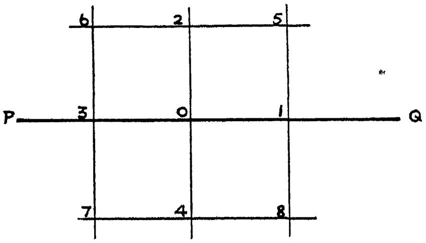


Figure 2

When O is on a boundary, say PQ, the points 2, 5 and 6, lying outside the field, are 'fictitious' and may be eliminated from (8) by means of the boundary conditions (7). At O, these are

$$\left.\begin{aligned}v_1-v_3+u_2-u_4 &= 0,\\ v_2-v_4+\sigma(u_1-u_3) &= 0,\end{aligned}\right\}$$

giving u_2, v_2 in terms of points on and internal to the boundary. Application of the boundary conditions at points 1 and 3 give u and v at 5 and 6 respectively. The results of elimination are

$$\left.\begin{aligned}\{2-\sigma(1+\sigma)\}(u_1+u_3)+2(1-\sigma)u_4-(1-\sigma)(v_1-v_3)-u_0\{6-2\sigma(2+\sigma)\} &= 0,\\ -2\sigma(v_1+v_3)+4v_4-2\sigma(u_1-u_3)-4(1-\sigma)v_0 &= 0,\end{aligned}\right\} \tag{9}$$

and similar equations hold for points on vertical boundaries. We are thus left with two equations for two unknowns, u and v, at every nodal point, including boundary points: and a solution can be obtained by relaxation in the usual way.

One-eighth of the side of the square was taken as the mesh-length, and u and v deduced at nodal points. By symmetry, only one-quarter of the plate required consideration, the conditions on Ox being $v=0$, $\partial u/\partial y=0$, and on Oy, $u=0$, $\partial v/\partial x=0$. The stress X_x was computed along Oy, and its distribution is given in curve III of figure 3.

Timoshenko used a strain-energy method, and obtained two successive approximations to this stress-distribution, the better one of which is shown in curve II of figure 3. Curve I shows the magnitude of the parabolic loading. Even with this coarse net, the agreement with Timoshenko's result is close.

4. When the boundaries are curved, the practical difficulties are greater, and simplifying assumptions have to be made. The problem considered (figure 4) was a circular plate of radius 5, with a concentric hole of radius 2. Equal radial displacements are imposed round the hole, and no traction acts on the outer boundary.

By symmetry, only one-eighth of the plate requires consideration. The governing equations are equations (3), and the boundary conditions are:

$$\left.\begin{array}{ll}\text{on } AB, & u=\cos\theta,\quad v=\sin\theta,\\ \text{on } CD, & 2\left(\dfrac{\partial u}{\partial x}+\sigma\dfrac{\partial v}{\partial y}\right)\cos\theta+(1-\sigma)\left(\dfrac{\partial v}{\partial x}+\dfrac{\partial u}{\partial y}\right)\sin\theta=0,\\ & 2\left(\dfrac{\partial v}{\partial y}+\sigma\dfrac{\partial u}{\partial x}\right)\sin\theta+(1-\sigma)\left(\dfrac{\partial v}{\partial x}+\dfrac{\partial u}{\partial y}\right)\cos\theta=0,\\ \text{on } BD, & v=0,\quad \dfrac{\partial u}{\partial y}=0,\\ \text{on } AC, & u=v.\end{array}\right\} \tag{10}$$

Figure 5 shows the first net investigated. Internal points whose finite-difference formulae involve external points are at A, B, C, D; points 1, 2, 3, 4 are fictitious, and P and Q lie on the boundary. Elimination of these fictitious points calls for some rather tedious computation, and approximations are used to reduce the labour without affecting the accuracy too greatly. No fixed rule can be stated, since the proper approximations depend largely on the position of the boundary with respect to the mesh lines. The following approximations were used in figure 5:

Regarding P as a fictitious point, then

$$2a\left(\frac{\partial u}{\partial x}\right)_P=3u_P-4u_A+u_E,\qquad 2a\left(\frac{\partial u}{\partial y}\right)_P=3u_P-4u_B+u_C.$$

Similar formulae for v derivatives and subsequent substitution in (10), give u_P, v_P in terms of internal points. The boundary conditions at L are satisfied by

$$a\left(\frac{\partial u}{\partial x}\right)_L\simeq u_3-u_B,\qquad 2a\left(\frac{\partial u}{\partial y}\right)_L\simeq 2a\left(\frac{\partial u}{\partial y}\right)_B=u_P-u_C,$$

with similar formulae for v derivatives. Substitution in (10) gives u_3, v_3 in terms of internal points.

At M we have

$$a\left(\frac{\partial u}{\partial x}\right)_M \simeq u_4 - u_C, \quad 2a\left(\frac{\partial u}{\partial y}\right)_M \simeq 2a\left(\frac{\partial u}{\partial y}\right)_4 = u_3 - u_Q,$$

and at Q,

$$2a\left(\frac{\partial u}{\partial x}\right)_Q = 3u_Q - 4u_D + u_G,$$

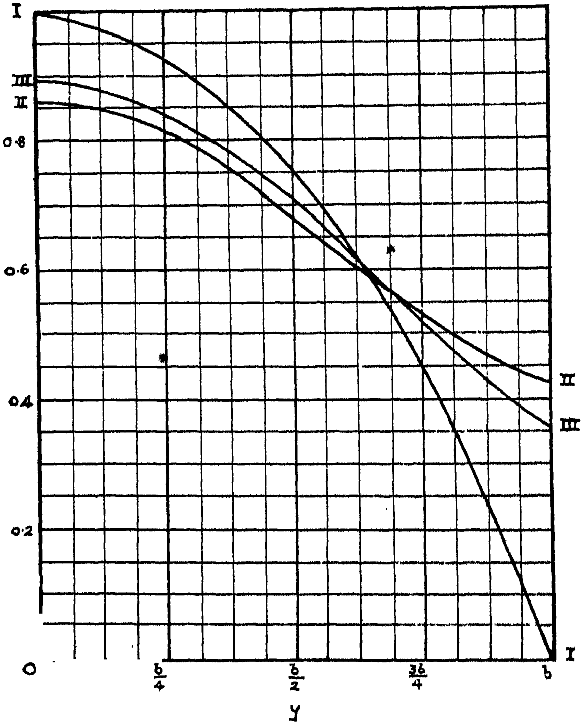


FIGURE 3

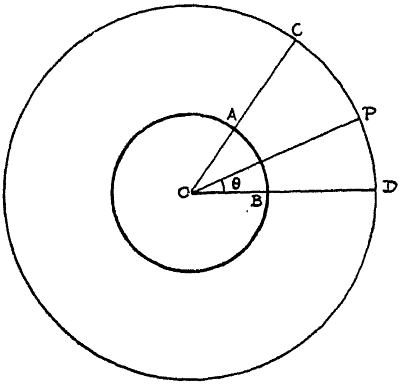


FIGURE 4

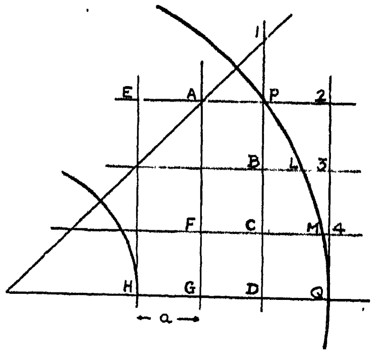


FIGURE 5

with similar formulae for v derivatives. Substitution in (10) gives u_4, v_4, u_Q, but the values at points 2 and 1 are still left undetermined, and for them we assume a parabolic continuation of u and v outside the field. The requisite type of formula is

$$u_2 = 3u_B - 3u_F + u_H.$$

Substitution of these fictitious values in the governing equation (8) gives two simultaneous equations at every internal point. (Points P and Q could have been regarded as internal points; but for them the finite-difference formulae involve points which are too far outside the field to permit accurate or easy elimination.)

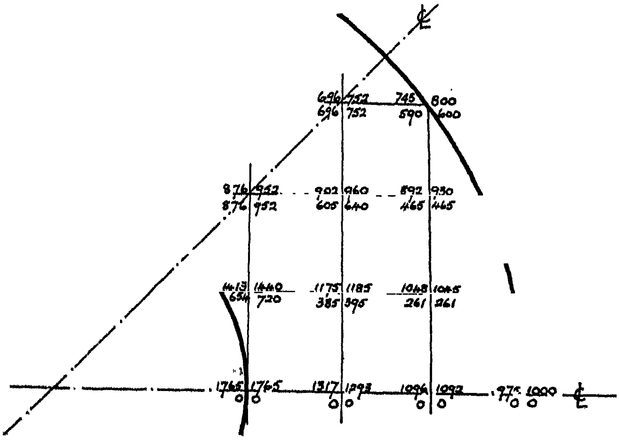


FIGURE 6

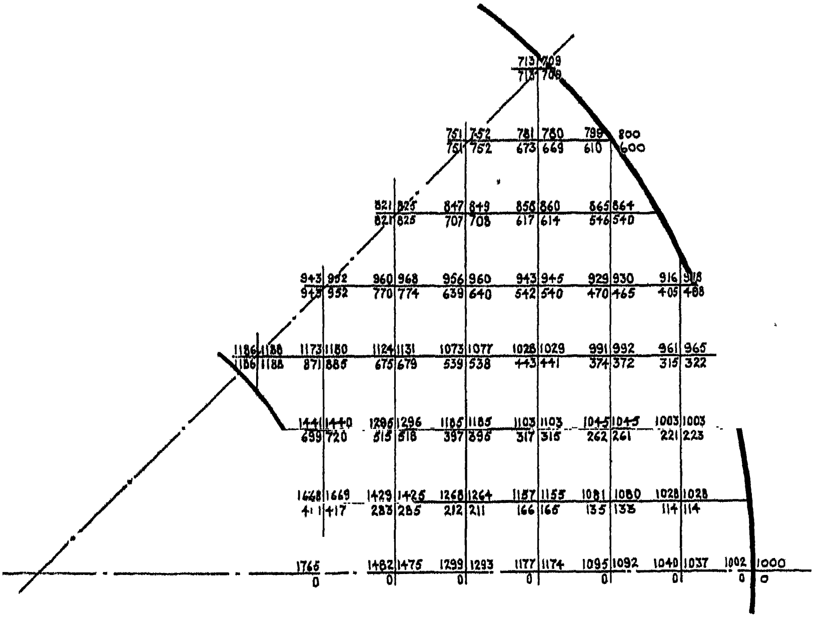


FIGURE 7

The resulting equations are complicated, but many of the coefficients are small enough to be neglected in relaxation, though they must of course be included in the calculation of residuals.

The results of relaxation on this net are given in figure 6, with analytical values, for comparison, to the right of the nodal points.

5. A finer net was investigated, with results which are given in figure 7; their closeness to the analytical solution, especially near the outer boundary, is very satisfactory. Near the inner boundary, on which u and v are given, the accuracy is not so good. The method adopted near this boundary was that of Part VII A, that is u and v were regarded as continuing linearly across the boundary. The accuracy would probably be improved by taking u, v to be parabolic across the boundary. In the general case in which a central point O is surrounded by eight points (figure 8) whose distances from O are a_r $(r = 1 \text{ to } 8)$, the finite-difference equivalents of equations (3) become

$$\frac{4}{a_1+a_3}\left(\frac{u_1}{a_1}+\frac{u_3}{a_3}\right)+\frac{2(1-\sigma)}{a_2+a_4}\left(\frac{u_2}{a_2}+\frac{u_4}{a_4}\right)+\frac{1+\sigma}{a_5+a_7}\left(\frac{v_5}{a_5}+\frac{v_7}{a_7}\right)-\frac{1+\sigma}{a_6+a_8}\left(\frac{v_6}{a_6}+\frac{v_8}{a_8}\right)$$
$$-u_0\left\{\frac{4}{a_1a_3}+\frac{2(1-\sigma)}{a_2a_4}\right\}+(1+\sigma)\,v_0\left(\frac{1}{a_6a_8}-\frac{1}{a_5a_7}\right)=0,$$

$$\frac{4}{a_2+a_4}\left(\frac{v_2}{a_2}+\frac{v_4}{a_4}\right)+\frac{2(1-\sigma)}{a_1+a_3}\left(\frac{v_1}{a_1}+\frac{v_3}{a_3}\right)+\frac{1+\sigma}{a_5+a_7}\left(\frac{u_5}{a_5}+\frac{u_7}{a_7}\right)-\frac{1+\sigma}{a_6+a_8}\left(\frac{u_6}{a_6}+\frac{u_8}{a_8}\right)$$
$$-v_0\left\{\frac{4}{a_2a_4}+\frac{2(1-\sigma)}{a_1a_3}\right\}+(1+\sigma)\,u_0\left(\frac{1}{a_6a_8}-\frac{1}{a_5a_7}\right)=0.$$

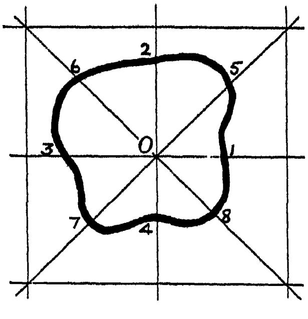


FIGURE 8.

The equations for points near the curved boundary are then obtained by putting a_r equal to the standard mesh length a or $a\sqrt{2}$ for those points lying inside the boundary, and to the measured distances for those points lying on the boundary, where the values of u and v are specified.

These formulae, though lengthy, are not more cumbersome than those already used at the outer boundary, and might have been applied with advantage; but this was not done, because the purpose of the investigation was to see how well the boundary conditions (10) could be satisfied and used in the relaxation process. The accuracy attainable would seem to be sufficient for all practical purposes.

Grateful acknowledgement is made to Dr J. R. Green, for collaboration in the devising of the method; to Mr D. N. de G. Allen, for the original suggestion, and to Miss G. Vaisey, for the computation of the second problem; to Dr R. V. Southwell, for overall supervision.

References

Fox, L. & Southwell, R. V. 1941 *Phil. Trans.* A, **239**, 419.
Timoshenko, S. 1934 *Theory of elasticity.* New York: McGraw Hill Co.

Vertical transport of heat by turbulence in the atmosphere

By C. H. B. Priestley and W. C. Swinbank

(*Communicated by E. Gold, F.R.S.—Received* 21 *June* 1946—*Read* 27 *February* 1947)

A necessary consequence of the classical theory of the turbulent transfer of heat in the atmosphere is that the flux of heat is in the direction from high to low potential temperature, and this normally involves the flux being from low to high actual temperature. On examination this is shown to be consistent with the second law of thermodynamics.

However, the theory assumes that an eddy may be regarded as a normal sample of the population at its level of origin; but when account is taken of buoyancy effects in a necessarily heterogeneous atmosphere it is clear that this assumption is incorrect. It also becomes apparent in the discussion that a definition of the concept 'level of origin' of an eddy more rigid than that given in the classical theory is necessary, and this has been given. It is shown that a component of heat flux, here called 'convective turbulence', must result from the buoyancy forces. This is always directed upwards and is thus opposed to the general downward transfer of mechanical turbulence.

Records of the fluctuations in temperature at fixed levels show amplitudes which cannot be explained by mechanical turbulence. The traces further indicate that convective turbulence may at times completely dominate the direction of the flow of heat, a conclusion which is supported by reference to some synoptic examples. This result helps to explain the normal increase of potential temperature with height.

Dimensional considerations provide a rough measure of the dependence of the buoyancy flux on the lapse rate and amplitude of fluctuations, and also allow the total flux to be written in the classical form $\rho c_p K'\left(\frac{\partial T}{\partial z}+\Gamma\right)$. K', which may be positive or negative, will depend upon the synoptic situation. Some of the implications in synpotic and general theoretical meteorology are discussed in detail. The turbulent transfer of physical quantities other than heat is briefly considered.

1. Introduction

The vertical transport of heat in the atmosphere is mainly effected by the following mechanisms: eddy or turbulent diffusion, the mass movement of air in subsidence, convexion and upslope motion at fronts, radiation and the transfer of latent heat of vaporization of water by the vertical interchange of water vapour. Transfer of heat by molecular diffusion can be regarded as negligible. This paper will be concerned primarily with the first mechanism, and the discussion will be restricted as far as possible to occasions when mass transfer of air and latent heat effects are not operative.

The chief agent in the interchange of radiant energy in the lower atmosphere is water vapour, the activity of which is all in the infra-red part of the spectrum. This interchange has been discussed by Brunt (1929), who concludes that the radiative transfer of heat in the lower atmosphere, while enormously greater than molecular transfer, can in general be neglected in comparison with turbulent transfer. Brunt's work was based on Hettner's values of the coefficients of absorption. Since recent work has shown that these coefficients are too high, it appears that water vapour should play a yet smaller part in its effect on the redistribution of temperature than was previously supposed.

Recent work in America on the water-vapour spectrum has been incorporated in an atmospheric radiation chart designed by Elsasser (1942). Work with this chart has shown that, at levels in the atmosphere away from the surface, radiative transfer is quite negligible compared with accepted values for the flux of heat by eddy diffusion. This conclusion is confirmed by Elsasser himself and Ashburn (1940).

The effect of radiation is to carry heat from high to low temperature and, as will be shown later, is thus in general opposed to the direction of transfer by eddy diffusion. But the evidence quoted above shows that, over short intervals at least, radiative transfer is completely outweighed by eddy transfer.

2. THE CLASSICAL THEORY OF EDDY DIFFUSION

The vertical transport of heat by eddies in the atmosphere has been discussed by Taylor (1915) in a treatment later modified by Brunt (1929, 1939). They considered the transfer of heat by eddies across a horizontal surface in the atmosphere at height z (strictly an isobaric surface should be considered, but the approximation is amply justifiable). The pressure, environmental temperature and density at this level are denoted by p, $T_0(z)$ and $\rho_0(z)$ respectively.

An eddy which crosses this level with upward vertical velocity w is regarded as having a density ρ and temperature

$$T_0(z) + T',$$

where T' may be termed the 'temperature anomaly'. The upward flux of heat across the surface per unit area per unit time is then

$$c_v(\overline{\rho w}\,T_0 + \overline{\rho w T'}),$$

c_v being the specific heat of air at constant volume. The bars denote mean values over the area. Since there can be no net transfer of mass across an isobaric surface, the average value of ρw is zero when the area and the time considered are both sufficiently great, and the flux of heat reduces to

$$c_v \overline{\rho w T'}.$$

In this expression the variation of density between eddies crossing the datum level has been allowed for. But whereas w and T' assume both positive and negative values, ρ is an essentially positive quantity which in practice varies by only a small percentage of its mean value. The existence of an appreciable flux depends ultimately on the existence of a correlation between w and T', and with sufficient accuracy the flux may be represented as

$$\rho_0 c_v \overline{wT'}. \tag{1}$$

Professor Taylor and Professor Cowling have pointed out, in correspondence, that this is not the only flux to be considered. There is also a mean rate of working $p\overline{w}$ when there exists a mean velocity $\overline{w}$ through the pressure surface. Writing ρ' for the 'density anomaly' defined in the same way as T', then in a constant-pressure surface

$$\frac{\rho'}{\rho_0}+\frac{T'}{T_0}=0.$$

Now $$\overline{\rho w}=\rho_0\overline{w}+\overline{\rho' w}=0,$$

so that $$\overline{w}=-\frac{1}{\rho_0}\overline{\rho' w}=\frac{1}{T_0}\overline{wT'},$$

and the flux of mechanical energy is

$$p\overline{w}=\frac{p}{T_0}\overline{wT'}=\rho_0(c_p-c_v)\,\overline{wT'}$$

(c_p is the specific heat of air at constant pressure). When the two fluxes are combined, the total flux of energy is accordingly

$$\rho_0 c_p \overline{wT'}. \tag{2}$$

The classical theory regards an eddy as having originated at a depth l below the surface, where the temperature is $T_0-l\frac{\partial T_0}{\partial z}$, $-\frac{\partial T_0}{\partial z}$ being the lapse rate of temperature, assumed constant in the layer z to $z-l$. The eddy is assumed to be an average sample of the air at its level of origin, and to move adiabatically during its transfer to the level z. On these assumptions the temperature anomaly T' is equal to $-l\left(\frac{\partial T_0}{\partial z}+\Gamma\right)$, where Γ is the dry-adiabatic lapse rate of temperature, and the upward flux of energy across the surface is then

$$F=-\rho_0 c_p \overline{wl}\left(\frac{\partial T_0}{\partial z}+\Gamma\right).$$

The product wl is supposed essentially positive for both upward and downward moving eddies, since w and l have the same sign for a given eddy. The mean value of this product $\overline{wl}$ is denoted by K, called the coefficient of eddy diffusivity. The flux of energy, written as

$$F = -\rho_0 c_p K\left(\frac{\partial T_0}{\partial z} + \Gamma\right), \tag{3}$$

is thus necessarily from high towards low values of potential temperature, since, if θ = potential temperature, $\frac{\partial \theta}{\partial z} = \frac{\theta}{T}\left(\frac{\partial T}{\partial z} + \Gamma\right)$.

3. SOME ASPECTS OF THE CLASSICAL THEORY

There are two aspects of the effects of the turbulent transfer of heat that can be examined in the light of recognized principles of thermodynamics and mechanics.

Equation (3) shows that the transfer of heat by eddy diffusion is from regions of high to regions of low potential temperature. A characteristic feature of upper-air temperature soundings is the universal decrease of temperature with height throughout the troposphere (except occasionally in the layers near the ground), and the stability of the air with respect to the dry-adiabatic lapse rate. It is readily seen that this latter effect implies increase of dry-bulb potential temperature with height. Thus the general effect of eddy diffusion throughout the troposphere, in the absence of evaporation and condensation processes, is to transfer heat downwards from regions of low to regions of higher temperature. This may at first sight appear to be a contravention of the well-established Second Law of Thermodynamics. The apparent difficulty can be resolved when it is realized that turbulent transfer in the free atmosphere derives from the existence of vertical wind shear; and wind shear is an effect of horizontal temperature gradients which are ultimately derived from heat which has been transferred upwards from regions of high to regions of low temperature. Thus the downward transfer of heat by eddy diffusion against the temperature gradient may be regarded as one stage in a process which, viewed as a whole, does not conflict with thermodynamical principles.

There is also an aspect of the effect of turbulent diffusion of heat which appears to conflict with accepted principles of mechanics, but which again can be resolved in terms of the wind shear. If one considers a layer of the atmosphere wherein the lapse rate of temperature is less than the dry-adiabatic, and excludes cases of condensation and evaporation, then the ultimate condition of the layer as a result of eddy transfer of heat will be one of potential temperature uniform with height. The gravitational potential energy of the layer in this state is greater than in the initial state, and this would appear to conflict with the minimum potential energy principle of mechanics. The work necessary to effect this increase in gravitational potential is done by the wind shearing stresses, and when a vertical wind shear exists it can be regarded as a form of potential energy. As shown by Petterssen & Swinbank (unpublished), the wind shear across a layer of uniform potential temperature is very small, and so the work done by the shearing stresses to increase the potential energy of the layer reduces the wind shear potential energy. Thus when all forms of potential energy are considered, the consequence of eddy transfer of heat is not necessarily an increase of total potential energy. Indeed Richardson (1920) has used as a principle that the work against gravitational forces in turbulent mixing is done by the shearing stresses.

On the other hand, it remains a problem to explain the almost universal increase of potential temperature with height in the free atmosphere. The classical theory of eddy transfer presents a powerful mechanism which constantly strives to destroy this increase. The theoretical determination of the time necessary to approach such a state (it would clearly take infinite time to achieve it completely) is difficult, but rough approximation shows that layers of a few thousand feet in depth, under the undisturbed influence of turbulent mixing, should approach the adiabatic state in a time of the order of magnitude of hours, assuming accepted values of K in the free atmosphere of order 5×10^4 c.g.s. (see Brunt 1939, p. 255 or Haurwitz 1941, p. 212). Occasions when large-scale vertical motion, and thereby condensation effects, can be excluded from consideration for considerable periods are fairly frequent. Thus a lapse rate approaching the adiabatic through appreciable depths of the troposphere should be a phenomenon of fairly frequent observation. But a feature of upper-air temperature soundings is the singular lack of such distributions of temperature away from the surface layers.

Certain other criticisms of the classical theory may be made on logical grounds. Equation (2) shows that a flux of energy can only result when the temperature at the level considered is subject to the fluctuations represented by T'. Thus when a flux of energy through the atmosphere is operative, the temperature at any level must fluctuate, both in time and in the horizontal plane. In spite of this necessary variation the classical theory, though purporting to assume that an eddy starts as a 'normal' sample of its environment, in fact regards it as an 'average' sample. 'Normal' implies randomness of selection at the level of origin so that a normal sample would be subject to the fluctuations appropriate to that level. These fluctuations are not allowed for in the construction of the equation of flux, and the consequences of this neglect will be considered in the next section.

A further difficulty lies in the use of the word 'originate'. Taylor (1915) describes an eddy as breaking away from the surrounding air, thereby implying that at the time when it can be considered a normal sample of its surroundings it is also (vertically) at rest relative to them. This assumption was later discarded when the possibility of different quantities having different diffusion coefficients was discussed in greater detail. If, however, it be admitted that an eddy is typical of its environment at different levels according to the property under consideration, then there is no reason to suppose that it is vertically at rest at any of these levels. It therefore does not follow that w, the upward component of vertical velocity, and l, the vertical distance travelled upwards since the eddy was last a normal specimen of its surroundings, are necessarily of the same sign. It is quite possible for an eddy to maintain its identity and to traverse the datum level successively in opposite directions without, at intermediate times, going through stages where it is typical of its surroundings with regard to each transferable property in turn. In fact there is no reason why it should not do so according to any theory which attributes no special significance to the element being instantaneously at rest in the vertical.

4. MODIFICATION OF THE THEORY

A modified formulation of the theory of flux will now be given in an attempt to overcome the logical difficulties of the previous section, by adhering to the eddy mixing length method of approach while giving to the concepts 'level of origin' and 'normal sample of surroundings' a more exact physical definition than has been given hitherto.

Proceeding from expression (2), consider the eddy as having last been at rest (in the vertical) at a level l below the datum level z. It follows immediately from this definition, by geometrical requirement independent of physical arguments, that w and l are of the same sign for every eddy, whatever the nature of the turbulent motion. When last at rest the eddy may be supposed to have had an excess temperature T'' over the average value $T_0(z-l)$ of its surroundings. The lifetime of an eddy, as in the classical theory, is assumed short in comparison with the time scale of any net average rise or fall of temperature at any level. Then, if the eddy moves adiabatically through the distance l,

$$T_0(z)+T' = T_0(z-l)+T''-\Gamma l,$$

and, assuming $\partial T_0/\partial z$ to be constant over the range of the mixing length,

$$T' = -l\left(\frac{\partial T_0}{\partial z}+\Gamma\right)+T''. \tag{4}$$

It should be noted that T' is greater than, equal to, or less than T'' according as the lapse rate is respectively greater than, equal to, or less than Γ.

Substituting from (4) in (2), the flux of energy across the datum surface is seen to be

$$F = \rho_0 c_p\left[-\overline{wl}\left(\frac{\partial T_0}{\partial z}+\Gamma\right)+\overline{wT''}\right]$$

which may be written $$F = \rho_0 c_p\left[-K\left(\frac{\partial T_0}{\partial z}+\Gamma\right)+\overline{wT''}\right], \tag{5}$$

where K is now an essentially positive quantity.

The first term in (5) represents a transfer of heat from regions of high to regions of low potential temperature and the rate of transfer is proportional to the gradient of $T+\Gamma z$. This is essentially the term derived by the earlier theories, though if K

is to be analyzed by means of its components a somewhat different interpretation than heretofore is to be given to the definition of l.

The second term is the result of the additional consideration that eddies starting from any level may have temperatures differing from the average value of their surroundings. In fact it represents a discrimination in the direction of motion of individual eddies, as opposed to the purely indiscriminate mixing given by the first term. It must be emphasized that it refers to the correlation between the upward velocity of the eddy as it crosses the datum level z and the excess temperature of the eddy *when it was last at rest*. Thus if there is any overall tendency, however slight, for the warmer eddies to start to rise, the sign of the second term becomes essentially positive and results in a systematic upward flux of heat, irrespective of the sign of $\partial\theta/\partial z$.

Having separated the flux into two terms in (5) it will be convenient to refer to them by distinctive names, and the terms 'mechanical turbulence' and 'convective turbulence' are suggested as appropriate, the word turbulence being retained since even in the convective form the motion is heterogeneous and the behaviour of individual particles cannot be specified. The second term in (5) will also be referred to as the 'buoyancy term'.

Taylor (unpublished) has recognized that convective forces and dynamical forces result in different laws for the heat flux. His treatment, which has been further applied by Sutton (unpublished), assumes a flux proportional to $\frac{\partial T_0}{\partial z}+\Gamma$ through a K which may take alternative analytical forms, but the existence of the term called the buoyancy term has not been recognized. After this paper was written it was pointed out that Ertel (1942) has given an equation similar to (5) above, but his main deductions and conclusions differ fundamentally from those here presented.

Since the buoyancy term $\overline{wT''}$ is not a measurable quantity, the importance of the modification to the theory can only be inferred. This may be done directly from special observations made in investigations of turbulence, or indirectly from synoptic data. The synoptic approach is essentially the more tentative, and therefore the less conclusive: but since it is the more important to the meteorologist, and has received unduly little attention in previous publications on this subject, it will be given first.

[*Note*. In comparing the two theories quantitatively, it will frequently be necessary to estimate a value of K. Where this has to be done the usual practice is followed of considering all features of wind, stability, and height above ground, and assigning the most appropriate value by reference to standard works on the subject. Specific reference is quoted in most cases. This procedure has the disadvantage that two theories are under comparison by means of a criterion which strictly applies only to one of them. But there appears to be no alternative method. Since the values quoted in the reference works are deduced by other applications of the classical theory, the procedure followed here at least represents a test of the consistency and ability of the classical theory to explain all the observed phenomena.]

5. CONSIDERATION OF SOME SYNOPTIC DATA

Appeal will here be made to three distinct types of synoptic data. In each case it will be shown that, whereas the classical theory fails to account adequately for certain observed facts, the additional consideration of convective turbulence will provide a qualitative explanation of the observed discrepancies.

(i) The problem of explaining the average temperature distribution in the troposphere has been referred to in § 3. Over a long period of time the action of convective turbulence will bring about an increase of potential temperature with height, and the rare occurrence of dry-adiabatic layers away from the surface requires no further explanation. This result is vital in cloud and rain forecasting, and it may be that the effect is of importance even if the normal fluctuations of temperature at, say, 10,000 ft. are found on investigation to be only of the order of a fraction of a degree. Looking a little further, it is possible that an approximate balance between mechanical and convective turbulence at very high levels may account for the ability of radiative processes to dominate the heat transfer, and so achieve the isothermal state characteristic of the lower stratosphere, whereas mechanical turbulence would cause a decrease of temperature with height.

(ii) In the case where a super-adiabatic lapse rate exists from the ground upwards, then in that region where the decrease of potential temperature with height gives way to an increase the classical theory would predict an accumulation of heat. If it is possible to determine two pressure levels at which the lapse rates throughout an interval of time remain respectively greater and less than the dry-adiabatic, it is possible to assess the amount of heat flowing into the region bounded by these levels.

Aircraft soundings are sufficiently detailed to allow identification of such levels, and for a 6 hr. time interval it is frequently possible to determine two such levels not more than 15 mb. apart. The gradient of potential temperature at either boundary of this layer is usually about 10^{-4}° C/cm. If for the lower level, where the lapse rate is super-adiabatic, one assumes (Brunt 1939; Haurwitz 1941) a K of 10^5 c.g.s. units, then heat will pass into the layer at a rate of about $2\cdot5\times10^{-3}$ cal./cm.^{2}/sec., which would cause a rate of increase of temperature of the layer of about 3° C/hr. Some inflow of heat through the upper boundary should yield a still greater rate of warming.

An examination was made of a year's aircraft soundings from Aldergrove near Lough Neagh in Northern Ireland, and all cases noted where a super-adiabatic lapse rate was sufficiently marked and deep to be immediately noticeable on a temperature entropy diagram. Of the 733 occasions when a super-adiabatic lapse rate persisted for 6 hr. or more, over 80 % occurred in the 50 mb. nearest to the ground with the minimum potential temperature at some height between 100 and 1500 ft. The observers (pilots of Coastal Command of the Royal Air Force) are instructed to report temperature singularities, especially in the lower layers, and the instrumental accuracy is such that it is reasonably sure that the potential temperature (θ) as computed is less than 1·0° F in error.*

A selection was then made of those cases in which the point of minimum potential temperature did not move by more than 15 mb. in 6 hr. By inspection of the synoptic charts, situations in which the temperature change due to horizontal advection was significant were eliminated. Of the remaining cases, only 54 % showed an increase of temperature in the 6 hr. at the level of minimum potential temperature, while 35 % showed a decrease. The mean increase of temperature over all cases was 0·6° F/6 hr., which is a very small fraction of that predicted above. One can only deduce either that the assumed value of K (which has been estimated from other applications of the same theory) is grossly in error, or that the heat brought upwards by turbulence through the unstable layer is capable of escaping through the upper boundary. Convective turbulence provides such a mechanism.

Figures from two situations are shown as examples of the working data:

date	1 June 1944		2 June 1944				5 June 1944		
time	1730	2330	0530	1130	1730	2324	0536	1130	1812
surface pressure (mb.)	1004	1008	1009	1009	1009	1010	991	995	998
pressure (mb.) at point (P) of minimum θ	965	970	980	975	970	975	955	950	950
temp. at P (°F)	48	43	43	49	50	47	45	45	45
condition immediately below P	S.A.	A.	S.A.	A.	S.A.	S.A.	S.A.	S.A.	S.A.

A. = adiabatic lapse; S.A. = super-adiabatic lapse, as identifiable on a temperature entropy diagram.

(iii) When the lapse rate is less than dry-adiabatic right down to the earth's surface, then the classical theory predicts a flux which is everywhere downwards. In considering the changes of temperature in the lowest layers of air and top inches of soil the heat quantities involved are then the eddy flux, radiation from the earth's surface, latent heat of vaporization of water condensed on or evaporated from the surface, insolation, radiation from clouds, and the change in sensible heat of the layers of ground. The number of factors involved, as is so often the case in meteorological processes, is so great that a quantitative consideration is difficult. Nevertheless this problem is so urgent and important in meteorology that it has been attempted for one type of situation in which many of the above quantities can be most accurately assessed.

The winter situation known as 'anticyclonic gloom' may persist for days and is characterized by extensive cloud of the inversion type at about 2000 ft., surface winds of 5 to 10 m.p.h., and a value of $\frac{\partial T_0}{\partial z}+\Gamma$ of about $5\times10^{-5\circ}$ C/cm. at 50 ft. This latter figure is derived from detailed lower-atmosphere soundings at Cardington, and confirms that given by Johnson & Heywood (1938) (table III, overcast days in December). The presence of the low cloud enables us to arrive at a reasonably sound estimate of the quantities of heat involved in radiation. If the cloud temperature exceeds that at the ground, which is sometimes the case in these situations, and both are assumed to radiate as black bodies, then water vapour between cloud and ground approximately halves the black-body radiation interchange, and at ordinary temperatures the ground will gain heat at a rate of about 0·1 cal./cm.^{2}/hr./° C temperature difference. This difference does not normally exceed 2° C. The estimation of the latent heat quantities is made difficult by the lack of experimental data, but information received from Rothamsted Experimental Station (for which we are grateful to Dr H. L. Penman) indicates that in these situations evaporation may be negligible, or may even be negative, i.e. water is condensed on the earth's surface.

---

* It is the practice in the Meteorological Office to measure temperature in °F.

The flux of heat through the 50 ft. level may be estimated from the classical theory by taking K as 2×10^4 (Best (1935), table XIII by inference; Johnson & Heywood (1938), tables XV and XXIII). The amount of heat flowing downwards in 2 days is then 36 cal./cm.^{2}. The heat capacity of 50 ft. of air can be neglected by comparison with that of a few inches of soil, and measurements of changes of soil temperature show that nearly all the heat should go to warming the top 4 in. This layer should then warm by 6° C. Records of earth temperature in the top 8 in. of soil taken over the past 2 years at Dunstable and Cardington show that in such situations the change in sensible heat of the top layers of the ground in 2 days is almost negligible. Thus, when consideration is given to the other heat quantities involved, it appears that in fact the flux of energy due to eddying motion must be a small fraction of the above 36 cal., and may even be directed upwards.

The three types of synoptic data thus strongly indicate that the classical theory fails to give a complete picture of the turbulent exchange. In each case the second term in equation (5) gives a term which acts in the right sense to account for the discrepancy.

6. VARIATIONS OF TEMPERATURE AT A FIXED LEVEL

A direct assessment of the importance of convective turbulence requires a knowledge of the magnitude of T''. This is not a measurable quantity, but its order of magnitude must be the same as that of T', the variation of temperature at a fixed level. Observations of T' have recently been taken at Cardington at a site over flat open grassland with good exposure. Temperature traces are obtained by means of a continuous record of the e.m.f. from sensitive thermocouples (36 s.w.g. copper, 40 s.w.g. Eureka). The cold junctions of these instruments are immersed in melting ice and the aspirated warm junctions, shielded from radiation, provide a continuous record of the air temperature at any required height, to which the instrument is hoisted by attachment to a balloon cable.

Corrections to the recorded amplitude for instruments of this type have been studied by Bilham (1935). For the instrument in use an initial difference between instrument and air temperature is reduced by 90 % in 5 sec., so that for fluctuations of period of the order of 10 sec. about one-half of the true amplitude is recorded.

Figure 1, (*a*), (*b*) and (*c*), shows temperature traces typical of certain synoptic situations. Time and temperature scales are indicated.

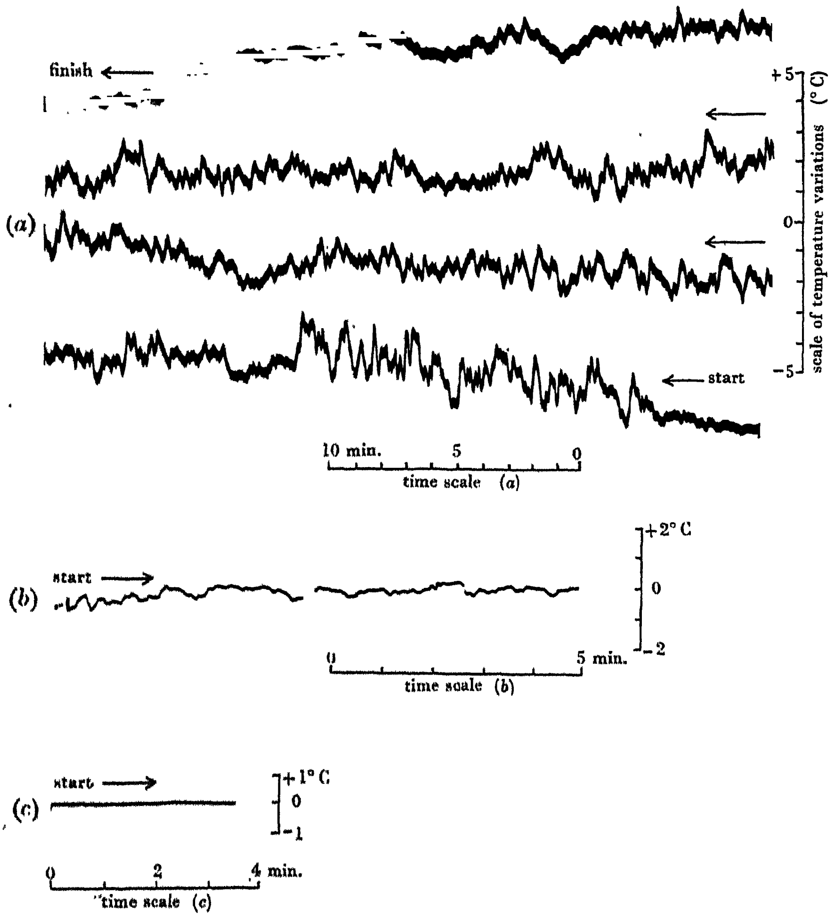


FIGURE 1. Variation of temperature recorded by a sensitive thermocouple thermograph at Cardington, Beds. (*a*) 4 ft. 6 in. above ground, 1450 to 1650 hr. G.M.T., 27 June 1945. Air in rear of cold front, northerly wind 6 to 10 m.p.h., variable amounts of cumuliform cloud, nine-tenths at start, clearing gradually to five-tenths. (*b*) 3700 ft., 1500 G.M.T. on 3 August 1945; clear, hot summer afternoon with little wind. (*c*) 3000 ft., 1400 G.M.T. on 11 September 1945; dull and overcast, with moderate southerly winds. Marked advective cooling of air.

Before discussing the full implications of these traces it is of interest to see to what extent their features can be explained by the classical theory of mechanical turbulence. If this mechanism is to control the eddy flux of heat, it must at the same time account for the distribution of temperature observed at any level. The upper limit to the fluctuation produced by this means is $l\left(\frac{\partial T_0}{\partial z}+\Gamma\right)$. In fact, as the ultimate mixing is a gradual rather than a discontinuous process, one would expect the fluctuations to be appreciably less than this.

Another upper limit, this time to the physically possible values of the mixing length, is set by the fact that heat is being continually diffused at the boundaries of the eddy. It is known that the 'size-spectrum' of atmospheric eddies contains eddies of all sizes down to very small dimensions. Since we are concerned in this section with hot and cold parcels of some tens of feet in diameter it is natural to suppose that the heat of such parcels is mainly diffused by eddies of smaller size which are, nevertheless, very active diffusing agents as compared with molecular processes. Scrase (1930) has shown that a large proportion of eddies at a height of 19 m. have periods of the order of 1 sec., and further that such eddies account for a value of K (for momentum) of 6×10^4 cm.^{2}/sec.

The equation for heat conduction with spherical symmetry may be written

$$\frac{\partial T}{\partial t} = K\left(\frac{\partial^2 T}{\partial r^2}+\frac{2}{r}\frac{\partial T}{\partial r}\right).$$

The solution of this equation which satisfies the initial conditions

$$T = T_0 \quad \text{for} \quad r \leqslant R, \qquad T = 0 \quad \text{for} \quad r > R$$

is given by

$$\frac{T}{T_0} = \tfrac{1}{2}[\phi(\xi_+)+\phi(\xi_-)]+\frac{1}{r}\sqrt{\frac{Kt}{\pi}}\,(e^{-\xi_+^2}-e^{-\xi_-^2}),$$

where

$$\xi_\pm = \frac{R\pm r}{\sqrt{(4Kt)}}$$

and

$$\phi(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,du.$$

If Scrase's value of K is adopted as being applicable to this scale of heat diffusivity, the solution shows that for an eddy of diameter 20 m. the excess temperature at the centre will fall to one-third of its initial value in about ¾ min. Half-way from the centre to the boundary the fall to one-third will occupy 5 sec. In practice the region outside the eddy boundary, or sink region, is subject to fluctuating temperatures which will further accelerate the loss of heat of the eddy.

Since w is of the order of 50 cm./sec., an eddy life of only some seconds will result in small values of l in the surface layers. It must be admitted that the magnitudes used in this calculation are no more than tentative, and that the small eddies may not be operative to the same extent at greater distances from the earth's surface. Nevertheless, it seems highly improbable that an l of some hundreds of metres, which would be necessary to cause the observed fluctuations, can occur.

Returning to the illustrated traces, figure 1 (a) shows the variations of temperature at 4 ft. 6 in. above the ground for a period of about 2 hr. (1450 to 1650 G.M.T. on 27 June 1945). The trace was taken a few hours after the passage of a cold front, with a light (6 to 10 m.p.h.) northerly wind and variable amounts of cumuliform cloud clearing from nine-tenths shortly after the start of the period. Allowing for Bilham's correction factor we may deduce representative (root mean square) values of T' of 1·5° C while the temperature was approximately steady, but 2° C and 0·5° C while the temperature was respectively rising and falling rapidly, with 0·5° C during the initial overcast period.

Although detailed information as to the temperature profile in the lowest layers was not available at Cardington on this occasion, reference to investigations by Johnson & Heywood (1938) and Best (1935) over similar sites permits a good estimate of the significant elements to be made. It is a fact of observation from Cardington that lapses greater than 4° C between screen level and 350 ft. have not been observed, so that the peaks and troughs of the temperature trace could only be due to eddies coming from very close to the surface and from several hundreds of feet respectively. On the basis of the computations made from Scrase's figures this explanation is most unlikely. (Note that the troughs are not noticeably broader than the peaks, as would then be the case.) Its improbability may be further demonstrated by watching drifting clouds of smoke near the ground, for it may be seen that only occasional 'puffs' rise or fall more than a few feet even in very turbulent conditions. Furthermore, Best (1935, table XIII) arrives at a value of K of about 10^3 at 4 ft. 6 in. on clear days, and with his representative value of w (table XXXIII) of 50 cm./sec. appropriate to winds of 6 to 10 m.p.h. one is not justified in assessing the value of l at more than about 1 m. Sutton (unpublished) deduces a yet much smaller value of K in these conditions. A lapse rate of 50Γ is a generous estimate for this level (Johnson & Heywood 1938, table III; Best 1935, table V). One may therefore quote 1° C as a very generous upper limit to the total range of fluctuations due to mechanical turbulence, ignoring the accelerating effect of the fluctuations in the sharing of heat between eddies and their surroundings. The actual range of 4° C cannot be explained on the basis of the transport of average eddies from different levels.

Figure 1 (*b*), obtained at a height of 3700 ft. about 1500 G.M.T. on 3 August 1945, is typical of the fluctuations on a clear hot summer afternoon with little wind. Representative values of T' from this trace are 0·5° C and the lapse rate was so close to the dry-adiabatic that a mixing length of 1500 ft. would be required to account for the fluctuations. Nothing is known of the values of w at this height, but an assumption of 100 cm./sec. would lead to an eddy-life of 8 min. and a value of K probably well in excess of 10^6, which can hardly be accepted. A smaller value of w, say 20 cm./sec., would imply an eddy-life of about 40 min., which must again be rejected on the basis of earlier calculations.

7. THE BUOYANCY TERM

It appears that mechanical turbulence, in failing to account for the observed fluctuations, must be equally inadequate to give a complete representation of the flux of heat. It remains to estimate whether the buoyancy term, i.e. the second term on the right-hand side of equation (5), is of significant magnitude compared with the first term. This estimation merely involves a rearrangement of the arguments of the previous section, using the same examples and figures.

It is not possible to measure T'' directly, but the appropriate values of l and $\partial T_0/\partial z$, when substituted in equation (4), show that T'' cannot differ in order of magnitude from T'. Observations of T' can therefore be used in assessing the magnitude of $\overline{wT''}$.

In example 1 (*a*) the estimated magnitudes are accordingly:

K	10^3 cm.^{2}/sec. or considerably less (Best 1935, table XIII).
$\dfrac{\partial T_0}{\partial z}+\Gamma$	0·005° C/cm.
$K\left(\dfrac{\partial T_0}{\partial z}+\Gamma\right)$	5° C × cm./sec. or much less.
T''	2° C.
w	50 cm./sec.

Thus even with the very generous magnitudes allowed for K and $\dfrac{\partial T_0}{\partial z}+\Gamma$ it requires only a very small correlation between w and T'' to produce a term of comparable magnitude, and the buoyancy flux may be several times greater than $K\left(\dfrac{\partial T_0}{\partial z}+\Gamma\right)$.

In example 1 (*b*), $\partial T_0/\partial z$ can be estimated from synoptic observations, and we have

K	10^5 cm.^{2}/sec. [a generous estimate for light winds, see Brunt (1939) and Haurwitz (1941)].
$\dfrac{\partial T_0}{\partial z}+\Gamma$	10^{-5}° C/cm.
$K\left(\dfrac{\partial T_0}{\partial z}+\Gamma\right)$	1° C × cm./sec.
T''	0·5° C.

w is not known, but it is clear that the buoyancy term is potentially the greater.

So far as absolute magnitudes are concerned, considerations of the heat balance in the lower layers indicate that the upward flux in example 1 (*a*) is of the order of 15 cal./cm.^{2}/hr. The flux from the term $K\left(\dfrac{\partial T_0}{\partial z}+\Gamma\right)$ is only 4 cal./cm.^{2}/hr., but a figure of 15 would be provided by the $\overline{wT''}$ term provided that the correlation coefficient between w and T'' were as high as 0·2. Such a correlation, which implies that only 4 % of the total variance Σw^2 is associated with variations in T'', is a very modest estimate. A similar correlation coefficient in 1 (*b*), assuming a representative w of 50 cm./sec., would give a flux at 3700 ft. of 4 cal./cm.^{2}/hr., which would represent a very important contribution to the heat balance.

Thus a wide variety of synoptic evidence points to the conclusion that the buoyancy term $\overline{wT''}$ is not only important but, in many situations, may be dominant in the heat flux. Away from the earth's surface the assumption that the magnitudes of T'' and T' are equal becomes more than ever justifiable, but w cannot readily be measured and so it is desirable to find some expression for the buoyancy flux which does not contain w explicitly. To do this further assumptions are necessary, and these may best be made from dimensional considerations. In the extreme case when the motion due to buoyancy predominates, an element which has moved upwards a distance z from its rest position will have an upward acceleration

$$\frac{dw}{dt} = w\frac{dw}{dz} \propto \frac{g}{T_0}\left[T'' \pm z\left(\frac{\partial T_0}{\partial z} + \Gamma\right)\right],$$

the sign being + or − according as the lapse rate is greater or less than the adiabatic, and the sign of proportionality being used to allow for the gradualness of the mixing process. From considerations of conduction in the previous section it may be supposed that the path L is short enough for the variations in T_0 and $\partial T_0/\partial t$ to be ignored when integrating, whence

$$w^2 \propto \frac{gL}{T_0}\left[T'' \pm \tfrac{1}{2}L\left(\frac{\partial T_0}{\partial z} + \Gamma\right)\right].$$

The path L will be determined by the magnitude of T'' and the lapse rate, and so will be assumed proportional to the length defined by

$$\frac{T''}{\left|\dfrac{\partial T_0}{\partial z} + \Gamma\right|}.$$

From this it follows that

$$w^2 \propto \frac{T''^2}{T_0\left|\dfrac{\partial T_0}{\partial z} + \Gamma\right|},$$

and the rate of upward flux due to buoyancy

$$\overline{wT''} \propto \frac{T''^2}{\sqrt{\left(T_0\left|\dfrac{\partial T_0}{\partial z} + \Gamma\right|\right)}}. \tag{6}$$

Formula (6) and its derivation are not applicable when the lapse rate is very close to the dry-adiabatic.

It is of some interest to see how the buoyancy term may be expected to behave in the region where the entities may be approximated by simple power laws in terms of the height. Reference to Johnson & Heywood (1938) figure 16, shows that on sunny June days the fluctuations fall off approximately as the square root of height, while from the first part of table III it appears that $\dfrac{\partial T_0}{\partial z} + \Gamma$ will fall off as some power between 1·5 and 2. Thus in such conditions the buoyancy term will fall off only very slowly with height. By measuring the temperature trends at different levels Sutton (unpublished) also deduces that the upward flux under these conditions shows only a very gradual decrease with height.

8. COMBINATION OF MECHANICAL AND BUOYANCY FLUX

In the preceding sections a sharp distinction has been drawn between mechanical and convective turbulence, as represented by the two terms in equation (5). This was done in order to emphasize the importance of the new term, but in reality the distinction is rather artificial. The two effects are necessary concomitants of each other, and are in many respects inseparable. This may be demonstrated by considering the extreme case in which buoyancy forces alone are operative, and in which each element is allowed to move without interference from other elements. If its motion is adiabatic, the element will then execute an approximately simple-harmonic motion with semi-range equal to the buoyancy length (defined below) and no net flux of heat will result. In order to account for a flux of heat, allowance must be made for the mutual interaction of the elements in the sharing of heat content and in interference with relative motions. It is known that mechanical turbulence will increase this mutual interaction many times over its magnitude under ordinary molecular processes; so that, even in cases where the flux of heat is predominantly controlled by convective turbulence, the mechanical turbulence plays an essential part in ensuring that heat carried upwards is shared by the upper layers of air and is not subsequently brought down again by a reversal of the buoyancy force on the same elements.

The mathematical link between the two terms can be illustrated by the introduction of a length

$$l'' = \frac{T''}{\dfrac{\partial T_0}{\partial z} + \Gamma}. \tag{7}$$

This length may be termed the 'buoyancy length' to distinguish it from the more familiar concept of mixing length. When the lapse rate is less than adiabatic, the buoyancy length has an immediate physical significance as representing the distance through which a buoyant element can move adiabatically before losing the net upward force upon it.

Substituting from (7) in (5), the rate of upward flux becomes

$$F = -\rho_0 c_p \overline{w(l-l'')} \left(\frac{\partial T_0}{\partial z} + \Gamma\right)$$

which may be written

$$F = -\rho_0 c_p K' \left(\frac{\partial T_0}{\partial z} + \Gamma\right), \tag{8}$$

an expression analogous to equation (3) for the classical theory. For lapse rates less than adiabatic, K' is capable of assuming both positive and negative values, so that the direction of flux is not necessarily determined by the gradient of potential temperature. K' will depend, among other things, on the amplitude of the temperature fluctuations at a fixed level.

Whereas, in preceding sections, examples and arguments have been adduced to show that the flux of heat may often be opposed to that indicated by the gradient of potential temperature, there is good evidence that in certain situations that gradient does in fact coincide with the direction of heat flux. For example, when air is being cooled by advection over cold ground, records of earth temperature at Cardington have shown that a usual rate of warming of the ground is a mean rise of temperature of 0·5° C/hr. in the top 8 in. of soil. This represents a very substantial rate of transfer of heat from air to ground and, if both mechanical and convective turbulence are operative in such conditions, it might be expected that the latter would be very small since its contribution to the heat flux is always upwards. Accordingly temperature traces have been made in the lowest few thousand feet on several such occasions. The trace shown at figure 1 (*c*), which is typical, shows the temperature behaviour at 3000 ft. on 11 September 1945, when marked advective cooling of the air was in progress, accompanied by overcast skies, and the top 8 in. of soil were being warmed at a mean rate of over 0·5° C/hr. Similar traces were obtained at 2, 4, 5 and 6000 ft.

Comparison of the trace with those shown in figure 1 (*a*) and (*b*) shows a notable absence of fluctuations and indicates that the upward flux of heat by convective turbulence is much less than in the examples studied in these figures, thus confirming the expectation above. Here, therefore, it seems that mechanical turbulence dominates the transfer of heat, and the gradient of potential temperature is in the direction of heat flow.

It is of interest to estimate the value of K' necessary to transport heat downwards fast enough to warm the earth at the rate indicated above. The heat capacity of the top layers of the soil at Cardington has been measured—it is approximately 0·6 cal./c.c. for the degree of wetness occurring on 11 September 1945. Thus, with $\frac{\partial T_0}{\partial z}+\Gamma$ approximately 10^{-4}° C/cm., we find K' to be about 5×10^4 c.g.s. units. This compares with Taylor's (1915) estimate of 3×10^3 over the sea (Great Banks) in a similar wind, but with a marked inversion replacing the isothermal condition at Cardington.

Thus in situations of advective cooling the flux of heat is directed downwards, K' being positive and of considerable magnitude. In fine weather in summer the lapse rate near the ground is greater than adiabatic and K' will be large and positive there, but may change sign in the region where the lapse rate becomes less than adiabatic with the flux of heat still directed upwards. When cold air is being warmed by advection the evidence points to a considerable flux of heat into regions, well away from the surface, where the lapse rate is stable, and thus indicates a large negative value of K'. This coefficient is then seen to depend very largely on the synoptic situation. The classical theory also stresses this dependence of K, but the considerations presented here indicate that not only the magnitude but also the sign of K' is a function of the synoptic situation.

9. GENERAL APPRECIATION AND EXTENSION TO OTHER PROPERTIES

In the extreme types of situation characterized by advective cooling and warming of air it seems that mechanical and convective turbulence respectively dominates the vertical heat flux. In the synoptic situations ranging between these extremes it seems likely that the one type of turbulence will gradually yield in importance to the other, and the various situations will be characterized according to the extent of predominance of the one or the other mechanism. It is possible that in certain 'quiet' types of weather a state of balance between the two may be approached. It is not possible to categorize, from this point of view, the different situations that occur without further extensive experimental investigation. This investigation should cover several heights at all times of day in a wide variety of synoptic situations, and should embrace different regions, with special attention to the nature of the underlying surface. At the same time an instrument might be devised to measure $\overline{wT'}$ directly, and thus the total flux as shown by expression (2).

From the theoretical standpoint there is the need for extension of the theory and observations to properties other than temperature. The arguments of § 5 applied to the quantity $T+\Gamma z$ are equally valid for any other physical property which may be considered as conserved in vertical motion up to the time at which an eddy loses its individuality by mixing processes. Examples of such properties are momentum in a given direction, specific humidity, dust content, and the salinity of the ocean. Now in a turbulent regime every property is likely to be subject at all levels to fluctuations similar to those represented by T''. Thus for each property whose transfer by turbulence is considered, the possibility of the existence of the discriminatory second term in (5) must be borne in mind. This applies immediately to any property which depends upon the density, or which affects the density, of the medium. In such cases the potential magnitude of the second term must, if possible, be estimated in relation to that of the first. From the fact that each property which evidences a discrimination is likely to do so to a different extent, it follows that the measure of transport may be expected to be different for various properties; and that a pronounced discrimination in the case of one property will affect the transport of a second property if the two are for any reason correlated. So far as is known this paper provides the first real physical justification for the use of different values of K for the transport by turbulence of different properties.

It also follows from expression (5) and the definition of l that properties which are subject to no discrimination in the direction of transport, and are uncorrelated with any property which is so subject, will be transported by the same mechanism, as represented by the first term, with the same value of K. This result depends on the validity of regarding the mixing process as discontinuous; which is justifiable provided that the mechanical disruption of the eddy is the predominant factor in mixing, as opposed to the gradual transfer of the property from one eddy to another by molecular processes.

In connexion with other properties it is of interest to note that a continuous record of the wet-bulb temperature was obtained synchronously with the dry-bulb trace of figure 1 (*a*). The similarity of the peaks and troughs in the two traces was extremely striking, but consideration of the implied fluctuations in humidity has been left for future work.

We wish to express our gratitude to the Director of the Meteorological Office for permission to publish this paper and for the use of observational material, and to members of the Meteorological Research Committee for criticisms of the first draft of the paper. Our thanks are also due to Mr H. Bateman and his staff at the Balloon Development Establishment at Cardington for assistance in the instrumental work, and for co-operation in the collection of data which have largely provided the material basis for this paper.

REFERENCES

Ashburn, O. M. 1940 *Bull. Amer. Met. Soc.* **21**, 205.
Best, A. C. 1935 *Geophys. Mem.* no. 65.
Bilham, E. G. 1935 *Quart. J. Roy. Met. Soc.* **61**, 259.
Brunt, D. 1929 *Proc. Roy. Soc.* A, **124**, 201.
Brunt, D. 1939 *Physical and dynamical meteorology*, 2nd ed. Camb. Univ. Press.
Elsasser, W. M. 1942 *Harv. Met. Stud.* no. 6.
Ertel, H. 1942 *Met. Z.* **59**, Heft 8, p. 250.
Haurwitz, B. 1941 *Dynamic meteorology.* New York: McGraw Hill.
Johnson, N. K. & Heywood, G. S. P. 1938 *Geophys. Mem.* no. 77.
Petterssen, S. & Swinbank, W. C. Large-scale turbulence in the free atmosphere (unpublished).
Richardson, L. F. 1920 *Proc. Roy. Soc.* A, **97**, 354.
Scrase, F. J. 1930 *Geophys. Mem.* no. 52.
Sutton, O. G. Convection in the atmosphere near the ground (unpublished).
Taylor, G. I. 1915 *Phil. Trans.* A, **215**, 1.
Taylor, G. I. Convection from a line source in a side wind (unpublished).

INDEX TO VOLUME 189 (A)

Acock, G. P., Garner, W. E., Milsted, J. & Willavoys, H. J. The dehydration of ammonium, potassium and some mixed alums, 508.
Alexandroff, P. On the dimension of normal spaces, 11.
Atmosphere. Vertical transport of heat by turbulence (Priestley & Swinbank), 543.
Atmospheric tides in the ionosphere. I. Solar tides in the F_2 region (Martyn), 241.

Biharmonic problems. Mixed boundary conditions in the relaxational treatment (plane strain or stress) (Fox), 535.
Blair, G. W. S., Veinoglou, B. C. & Caffyn, J. E. Limitations of the Newtonian-time scale in relation to non-equilibrium rheological states and a theory of quasi-properties, 69.
Bleaney, B. & Penrose, R. P. The inversion spectrum of ammonia at centimetre wavelengths, 358.
Bristow, J. R. Kinetic boundary friction, 88.
Buckingham, R. A. & Corner, J. Tables of second virial and low-pressure Joule-Thomson coefficients for intermolecular potentials with exponential repulsion, 118.
Bunn, C. W. & Garner, E. V. The crystal structures of two polyamides ('nylons'), 39.
Burnett, G. M. & Melville, H. W. Determination of the velocity coefficients for polymerization processes. I. The direct photopolymerization of vinyl acetate. II. The photosensitized polymerization of vinyl acetate by benzoyl peroxide. III. Effect of environment on polymerization processes, 456, 481, 494.

Caffyn, J. E. *See* Blair, Veinoglou & Caffyn.
Corner, J. *See* Buckingham & Corner.
Crawford, B. H. Visual adaptation in relation to brief conditioning stimuli [Abstract], 284.
Crystal structure of barium titanium. Temperature changes (Megaw), 261.
Crystal structures of two polyamides ('nylons') (Bunn & Garner), 39.

Dehydration of ammonium, potassium and some mixed alums (Acock, Garner, Milsted & Willavoys), 508.
Determination of the velocity coefficients for polymerization processes. I, II, and III (Burnett & Melville), 456, 481, 494.
Dimension of normal spaces (Alexandroff), 11.

Effect of crossed electric magnetic fields on the helium spectrum (Foster & Pounder), 287.
Environment. Effect on polymerization processes (Burnett & Melville), 494.
Explosives, rate of detonation (Jones), 415.

Fine structure of the lines of the H_2 spectrum which go down to the $2p^3\Pi$ states (Foster & Richardson), 175.
Fine structure of the spectral lines of H_2: measurements with a reflexion echelon (Foster & Richardson), 149.
Foster, E. W. & Richardson, Sir Owen. The fine structure of the lines of the H_2 spectrum which go down to the $2p^3\Pi$ states, 175.

Foster, E. W. & Richardson, Sir Owen. The fine structure of the spectral lines of H_2: measurements with a reflexion echelon, 149.
Foster, J. S. & Pounder, E. R. Effect of crossed electric magnetic fields on the helium spectrum, 287.
Fox, L. Mixed boundary conditions in the relaxational treatment of biharmonic problems (plane strain or stress), 535.
Fragmentation of shell cases (Mott), 300.

Garner, E. V. *See* Bunn & Garner.
Garner, W. E. *See* Acock, Garner, Milsted & Willavoys.
Gases after the electric discharge has passed. The surprising amount of energy which can be collected from (Rayleigh), 296.
Gaydon, A. G. & Whittingham, G. The spectra of flames containing oxides of sulphur, 313.
General kinetic theory of liquids. II. Equilibrium properties (Green), 103.
Gold. Slow neutron-induced activity (Siegbahn), 527.
Green, H. S. A general kinetic theory of liquids. II. Equilibrium properties, 103.
Grew, K. E. Thermal diffusion in mixtures of the inert gases, 402.

Harish-Chandra. Infinite irreducible representations of the Lorentz group, 372.
Hartley, Sir Harold, Antoine Laurent Lavoisier, 427.
Helium spectrum, effect of crossed electric magnetic fields (Foster & Pounder), 287.

Inert gases. Thermal diffusion in mixtures of the (Grew), 402.
Infinite irreducible representations of the Lorentz group (Harish-Chandra), 372.
Ionosphere, atmospheric tides (Martyn), 241.
Ionosphere, radiative equilibrium (Woolley), 218.
Interaction of steam and charcoal at low pressures (Strickland-Constable), 1.
Inversion spectrum of ammonia at centimetre wave-lengths (Bleaney & Penrose), 358.

Jones, H. A theory of the dependence of the rate of detonation of solid explosives on the diameter of the charge, 415.

Kinetic boundary friction (Bristow), 88.

Lavoisier, Antoine, Laurent (Hartley), 427.
Limitations of the Newtonian time scale in relation to non-equilibrium rheological states and a theory of quasi-properties (Blair, Veinoglou & Caffyn), 69.
Liquids. A general kinetic theory. II. Equilibrium properties (Green), 103.
Lorentz group. Infinite irreducible representations of the (Harish-Chandra), 372.

Martyn, D. F. Atmospheric tides in the ionosphere. I. Solar tides in the F_2 region, 241.
Megaw, H. D. Temperature changes in the crystal structure of barium titanium oxide, 261.
Melville, H. W. *See* Burnett & Melville.
Merton, Sir Thomas. On a method of increasing contrast in microscopy, 309.
Method of increasing contrast in microscopy (Merton), 309.
Microscopy. On a method of increasing contrast in (Merton), 309.
Milsted, J. *See* Acock, Garner, Milsted & Willavoys.
Mixed boundary conditions in the relaxational treatment of biharmonic problems (plane strain or stress) (Fox), 535.
Mott, N. F. Fragmentation of shell cases, 300.

Newtonian time scale. Limitations in relation to non-equilibrium rheological states and a theory of quasi-properties (Blair, Veinoglou & Caffyn), 69.

Oblique reflexion of very long wireless waves from the ionosphere (Wilkes), 130.

Penrose, R. P. *See* Bleaney & Penrose.
Photopolymerization of vinyl acetate (Burnett & Melville), 456.
Photosensitized polymerization of vinyl acetate by benzoyl peroxide (Burnett & Melville), 481.
Polyamides. Crystal structure (Bunn & Garner), 39.
Polymerization processes. Determination of the velocity coefficients. I, II and III (Burnett & Melville), 456, 481, 494.
Pounder, E. R. *See* Foster & Pounder.
Priestley, C. H. B. & Swinbank, W. C. Vertical transport of heat by turbulence in the atmosphere, 543.

Radiative equilibrium in the ionosphere (Woolley), 218.
Rayleigh, Lord. The surprising amount of energy which can be collected from gases after the electric discharge has passed, 296.
Relaxational treatment of biharmonic problems (plane strain or stress). Mixed boundary conditions (Fox), 535.
Richardson, Sir Owen. *See* Foster & Richardson.

Shell cases. Fragmentation of (Mott), 300.
Siegbahn, K. Slow neutron-induced activity in gold, 527.
Simple group of order 25920 (Todd), 326.
Slow neutron-induced activity in gold (Siegbahn), 527.
Solar tides in the F_2 region. Atmospheric tides in the ionosphere. I. (Martyn), 241.
Spectra of flames containing oxides of sulphur (Gaydon & Whittingham), 313.
Steam and charcoal. Interaction at low pressures (Strickland-Constable), 1.
Strickland-Constable, R. F. Interaction of steam and charcoal at low pressures, 1.
Surprising amount of energy which can be collected from gases after the electric discharge has passed (Rayleigh), 296.
Swinbank, W. C. *See* Priestley & Swinbank.

Tables of second virial and low pressure Joule-Thomson coefficients for intermolecular potentials with exponential repulsion (Buckingham & Corner), 118.
Temperature changes in the crystal structure of barium titanium oxide (Megaw), 261.
Theory of the dependence of the rate of detonation of solid explosives on the diameter of the charge (Jones), 415.
Thermal diffusion in mixtures of the inert gases (Grew), 402.
Todd, J. A. On the simple group of order 25920, 326.
Turbulence in the atmosphere. Vertical transport of heat by (Priestley & Swinbank), 435.

Veinoglou, B. C. *See* Blair, Veinoglou & Caffyn.
Velocity coefficients for polymerization processes. Determination I, II and III (Burnett & Melville), 456, 481, 494.
Vertical transport of heat by turbulence in the atmosphere (Priestley & Swinbank), 543.
Vinyl acetate. The direct photopolymerization (Burnett & Melville), 456.

Vinyl acetate. The photosensitized polymerization by benzoyl peroxide (Burnett & Melville), 481.
Visual adaptation in relation to brief conditioning stimuli [Abstract] (Crawford), 284.

Whittingham, G. *See* Gaydon & Whittingham.
Wilkes, M. V. The oblique reflexion of very long wireless waves from the ionosphere, 130.
Willavoys, H. J. *See* Acock, Garner, Milsted & Willavoys.
Wireless Waves. Oblique reflexion of waves from the ionosphere (Wilkes), 130.
Woolley, R. v. d. R. Radiative equilibrium in the ionosphere, 218.

END OF THE ONE HUNDRED AND EIGHTY-NINTH VOLUME (SERIES A)

I.A.R.I. 75

INDIAN AGRICULTURAL RESEARCH INSTITUTE LIBRARY, NEW DELHI.

Date of Issue	Date of Issue	Date of Issue
11.5.40		

GIPNLK—H-40 I.A.R.I.—29-4-55—15,000